PhD Degree in Molecular Medicine
European School of Molecular Medicine (SEMM)
University of Milan and University of Naples "Federico II"
Faculty of Medicine
Settore Disciplinare: MED/04

Characterization of genetic and epigenetic modifications in a model of inflammation-driven cancer

Agnese Collino
IFOM-EO Campus, Milan
Matricola n. R08890

Supervisor: Dr Gioacchino Natoli
IFOM-IEO Campus, Milan

Anno Accademico 2012-2013

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

γGT: gamma-glutamyltranspeptidase
Abcb4: adenosine triphosphate-binding cassette, sub-family B, member 4
AFP: α-fetoprotein
AID: activation-induced deaminase
ALT: alanine aminotransferase
AP1: activator protein 1
APOBEC: apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like
ARID: AT-rich interactive domain
ASBT: apical sodium bile salt transporter
ASCAT: allele-specific copy number analysis of tumors
ATF: activating transcription factor
BAZ2B: bromodomain adjacent to zinc finger domain, 2B
BCL6: B-cell CLL/lymphoma 6
BPTF: bromodomain PHD finger transcription factor
BRCA1: breast cancer 1
BRD8: bromodomain-containing protein 8
BRE: brain and reproductive organ-expressed
BSA: bovine serum albumin
BSEP: bile salt export pump
CA: cholic acid
CAGE: cap analysis gene expression
CCR: CC-chemokine receptor
CD3: cluster of differentiation 3
CDCA: chenodeoxycholic acid
CDKN2A: cyclin-dependent kinase inhibitor 2A
CFS: common fragile site
CGI: CpG island
CHFR: checkpoint with forkhead and ring finger domains, E3 ubiquitin protein ligase
ChIPseq: chromatin immunoprecipitation followed by new generation sequencing
CIMP: CpG island methylator phenotype
CNV: copy number variation
CTNBB1: catenin beta 1
CXCR: CXC-chemokine receptor
Cyld: cylindromatosis
Cyp27a: cytochrome P450, family 27, subfamily A
DCA: deoxycholic acid
DEN: diethylnitrosamine
DMSO: dimethyl sulfoxide
DNMT: DNA methyltransferase
DR4: death receptor 4
EDTA: Ethylenediaminetetraacetic acid
ETS: erythroblast transformation-specific transcription factor
EZH2: enhancer of zeste homolog 2
FACS: fluorescence-activated cell sorting
FADD: Fas-associated death domain
FANCF: Fanconi anemia, complementation group F
FDR: false discovery rate
FFPE: formalin-fixed paraffin-embedded
FIC1: familial intrahepatic cholestasis 1
FIMO: find individual motif occurrence
FISH: fluorescence *in situ* hybridization
Fos: FBJ murine osteosarcoma viral oncogene homolog
FVB: Friend virus B-type
FXR: farnesoid X receptor
G4: guanine quadruplex
GREAT: genomic regions enrichment of annotations tool
GO: gene ontology
Gp130: glycoprotein 130

H3K4me1: histone H3 lysine 4 mono-methylation
H3K4me3: histone H3 lysine 4 tri-methylation
H3K9me3: histone H3 lysine 9 tri-methylation
H3K27Ac: histone H3 lysine 27 acetylation
H3K27me3: histone H3 lysine 27 tri-methylation
H4K16Ac: histone H4 lysine 16 acetylation
H4K20me3: histone H4 lysine 20 tri-methylation
HBV: hepatitis B virus
HCC: hepatocellular carcinoma
HCV: hepatitis C virus
HDAC: histone deacetylase
HDM: histone demethylase
HDV: hepatitis D virus
HEPES: (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HMT: histone methyltransferase
HPV: human papillomavirus
IGF2: insulin-like growth factor 2
IL: interleukin
IFN: interferon
JDP: Jun-dimerizing partner
JMJD3: jumanji domain containing 3
JNK: cJun N-terminal kinase
KO: knockout
LOI: loss of imprinting
MACS: Model-based analysis of ChIPseq
MAPK: mitogen-activated protein kinase
Mdr2: multidrug resistance protein 2
MEN1: multiple endocrine neoplasia 1
MGI: mouse genome informatics
MGMT: O-6-methylguanine-DNA methyltransferase
MLH1: MutL homolog 1, colon cancer, nonpolyposis type 2
MLL: mixed-lineage leukemia
Myc: myelocytomatosis viral oncogene homolog
MyD88: Myeloid differentiation primary response 88
N.A.: not applicable
NFkB: nuclear factor kappa-light-chain-enhancer of activated B cells
NGS: next-generation sequencing
PBS: phosphate-buffered saline
PC: phosphatidylcholine
PCA: principal component analysis
PE: paired end
PEG: polyethylene glycol
PFIC: progressive familial intrahepatic cholestasis
PRC2: Polycomb repressive complex 2
PS: phosphatidylserine
qPCR: quantitative polymerase chain reaction
RNAi: RNA interference
RNS: reactive nitrogen species
ROS: reactive oxygen species
SCD1: Stearoyl-CoA desaturase 1
siRNA: small interfering RNA
shRNA: short hairpin RNA
SMAD: mothers against decapentaplegic homolog
SMARCB1: SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily b, member 1
SNV: single nucleotide variant
SPGP: sister of P-glycoprotein
SPI1: spleen focus forming virus (SFFV) proviral integration oncogene 1
SSC: saline-sodium citrate
STAT: signal transducer and activator of transcription
SWI/SNF: SWItch/Sucrose NonFermentable

TAK1: TGFβ activated kinase 1
TBS: Tris-buffered saline
TE: target enrichment
TERT: telomerase reverse transcriptase
TF: transcription factor
TGF: tumor growth factor
TNF: tumor necrosis factor
TRAIL: TNF-related apoptosis-inducing ligand
TRADD: TNF receptor-associated death domain protein
TRAF2: TNF receptor-associated factor 2
TRE: TPA (tetradecanoylphorbol acetate)-DNA responsive element
TRL: Toll-like receptor
TSS: transcription start site
TYK2: tyrosine kinase 2
UDCA: ursodeoxycholic acid
UTX: ubiquitously transcribed tetratricopeptide repeat, X chromosome
VHL: Von Hippel Lindau protein
WES: whole-exome sequencing
WGS: whole-genome sequencing
WRN: Werner syndrome ATP-dependent helicase
WT: wild type

FIGURES INDEX

TABLES INDEX

ABSTRACT

Chronic inflammation is causally associated to many types of tumor, and has been recently acknowledged as a cancer hallmark. Nevertheless, whether inflammation has an intrinsic mutagenic potential is still not directly proven or understood from a mechanistic point of view. Furthermore, it is as yet unclear whether inflammation could induce epigenetic modifications, and if these changes are relevant to tumor generation.

Therefore, the aim of this work was to assess inflammation-derived genomic and epigenomic modifications at multiple stages of tumorigenesis in *Mdr2*-knockout mice, a model of purely inflammatory hepatocellular carcinoma (HCC).

By ChIPseq profiling of H3K27Ac mark we reported the establishment of an inflammatory program specifically in hepatocytes starting from the pre-malignant, chronic inflammatory step. This inflammatory signature is retained up to the more advanced HCC stage, and is accompanied by the activation of members of the AP1 transcription factor family.

In parallel, by whole exome sequencing, we observed a high frequency of copy number amplifications and a very low number of point mutations in HCC nodules. Copy number variations occurrence was directly correlated to the grade of malignancy in each lesion. The JNK pathway was shown to be pervasively targeted by gene amplification, and to be involved in the adenoma-to-carcinoma transition. A comparable genetic landscape has been observed in a human liver cancer with similar etiology.

In conclusion, this study shows that in a model of inflammation-driven cancer an epigenetic inflammatory signature is early acquired and maintained throughout disease progression. On the contrary, genetic alterations appear only at later stages and mainly target the JNK pathway. Future dataset integration will help clarifying chronological relationship and possible mutual interplay between mutations and epigenetic changes.

INTRODUCTION

1. Genetic and epigenetic factors in cancer

1.1 Genetic alterations in cancer

The intuition of cancer as a "disease of the genome" came from Theodor Boveri at the beginning of the twentieth century [1]. Eighty years later, a proof of principle for this concept was eventually provided by the earliest reports of mutated cancer-causing genes [2, 3]. Classification of such genes as oncogenes and tumor suppressors became soon established [4]. The variety of genetic alterations occurring (such as point mutations, copy number variations, chromosomal rearrangements) began to be described, and to shed light on the underlying genomic complexity of cancer, on the multiplicity of tumor-causing genes and their variability across and within cancer types [5, 6].

Nevertheless, the field of "cancer genomics", namely the systematic and genome scale analysis of cancer genomes to identify recurrent alterations for specific tumor types, started off only recently. After a pilot project, the Cancer Genome Atlas was launched by the National Cancer Institute in US in 2009, in parallel to the establishment of an International Genome Consortium [7]. The challenge of creating a compendium of genomic alterations for each given tumor type was made easier by the advent of next-generation sequencing (NGS) approaches based on massively parallel sequencing [8]. From the first whole cancer genome study performed by NGS [9], where a single sequencing run could analyze around 1 billion bases, the technology evolved to the current throughput of >600 gigabases per run. As the costs of NGS progressively dropped, sequencing of large number of samples became more affordable, giving rise to a large diffusion of large scale studies carried out on primary patients samples.

An optimization of the NGS technology for the analysis of cancer genomes was dependent on the design of hybridization-based selection techniques (collectively referred to as "target enrichment") to sequence sub-portions of genome, specifically the whole complement of coding and non-coding exons (exome) [10]. Moreover, the development of robust bioinformatic approaches allowed taking advantage of NGS-

based platforms for the analysis of a panel of different genetic alterations (point mutations, copy number variations), or functional outcomes (transcriptomic measurements, epigenetic modifications profiling, chromatin structure analysis).

The explosion of sequencing data originated from NGS diffusion allowed providing a number of unprecedented insights on the variegated mutational spectra across different tumor types. The mutation frequency is highly variable, ranging from 0.5/Mb for acute myeloid leukemia to around 90/Mb in solid tumors such as melanoma [11, 12]. New evidences of the high dependence of mutational patterns on etiological factors (being them external, such as smoke or UV exposure, or internal, as DNA repair mechanism defects) were presented [13, 14]. Furthermore, new mutational processes have been uncovered. For example, a new regional hypermutation mechanism characterized by multiple base mutations occurring in *cis* close to rearrangement breakpoints, initially identified in breast cancer and soon after recognized also in many other tumor types, has been termed "kataegis" [15, 16]. This phenomenon is thought to depend on the activation-induced deaminase (AID) and apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like (APOBEC) families of proteins. More extensive damage comes from other new mutational mechanisms, consisting of catastrophic events (defined as one-step massive genomic alterations with respect to step-wise multiple mutations) resulting in high degrees of rearrangements. "Chromothripsis", or chromosomal shattering, occurs in ~2-3% of cancers and usually involves massive rearrangements of one or two chromosomes, with large portions of chromosomes fluctuating between two distinct copy number states [17]. This appears to result from errors in mitotic chromosomal segregation, which would cause the confinement of single chromosomes into "micronuclei". Micronuclear chromosomes would be subject to premature chromosome condensation and consequent breaks in the incompletely replicated DNA (chromosome pulverization): a few chromosomes might survive this process *via* aberrant non-homologous end joining, giving rise to densely rearranged elements [18].

Besides shedding light on newly identified mutational processes, high throughput cancer sequencing studies strengthened our knowledge on mechanisms already established as characteristic of tumoral conditions. Aneuploidy (an abnormal number

of chromosomes), a typical feature of cancerous cells, was systematically assessed on large-scale studies, along with focal copy number variations of variable entity. In particular, it has been estimated that large-scale events affect on average ~25% of a cancer genome, while 10% is target of focal gains/losses involving a median of 6-7 genes [19].

The huge amount of information regarding mutations and copy number alterations has been integrated in the search for conserved cancer signatures. The analysis of a catalogue of somatic mutations in >7,000 samples across the spectrum of human cancers yielded 21 distinct mutational signatures, some of which are conserved among different tumors while others are specific to a given type [16]. Although most cancer classes present at least two signatures, some cancers (such as liver, uterus and stomach) can display up to six signatures, suggesting a higher degree of complexity in their mutational processes [16]. A recent work on ~3,000 samples (from 12 different cancer types) retrieved 30 tissue-independent tumor subclasses, based on oncogenic signatures [20]. Interestingly, the same study observed that tumors, independently of their origin, present an inverse correlation between the number of somatic mutations and the number of recurrent copy number variations. Accordingly, tumors have been reported to subdivide into two main classes, one principally bearing somatic mutations and one predominantly characterized by copy number alterations [20].

Such efforts towards the integration of large amounts of genomic data will help in understanding tumor biology in terms of signatures systematically derived from functional alterations, and will possibly catalyze the elaboration of personalized cancer therapies.

1.2 Cancer epigenetics

Beside proposing the genetic origin of cancer, Boveri was probably also the first to observe an epigenetic aberration in cancer, by describing the abnormal appearance of chromatin in tumor cells as early as in 1929 [1]. This occurred more than a decade before the term "epigenetics" had been used for the first time by Conrad Waddington to define "the interactions of genes with their environment which bring the phenotype

into being" [21]. Epigenetics was originally defined as the study of changes in gene expression caused by mechanisms different from DNA sequence modifications, such as changes in DNA methylation and chromatin remodeling. The concept became further extended after the recent spread of genomic studies, and the field of "epigenomics", intended as the study of the "effects of chromatin structure, including the higher order of chromatin folding and attachment to the nuclear matrix, packaging of DNA around nucleosomes, covalent modifications of histone tails (acetylation, methylation, phosphorylation, ubiquitination), and DNA methylation" [22].

The cellular transformation process needs the acquired malignant features to be permanently encoded in order to allow their transmission and accumulation along the clonal expansion. Genetic mutations represent the most obvious mechanism to convey stable phenotypic changes. For this reason, cancer has been considered for a long time from a gene-centric viewpoint, with DNA mutations being assigned a primary role in generating its hallmark properties. This diverted the attention from the possible importance of epigenomic dysregulations, initially regarded as "surrogates" of genetic changes [23], and from their capability of being flexible, but still inheritable through multiple cell divisions.

The first demonstration of a connection between cancer and epigenetic processes came in 1983 from Feinberg and Vogelstein, who first described the hypomethylation of colorectal tumor cells DNA [24]. This observation was soon extended to a variety of tumors and pre-malignant lesions [25], DNA hypomethylation becoming an ubiquitous feature of tumor cells. Classical observation include a global hypomethylation of cancer cells DNA, occurring during early tumorigenic steps and thought to favor genomic instability [26], and focal hypermethylation at particular CG dinucleotide-rich regions called CpG islands (CGIs) [27]. CGIs have been found enriched in more than 70% of human promoters, and for this reason CGI hypermethylation is associated to transcription silencing [28]. More recently, a number of reports showed that most methylation differences between tissues (including normal *vs.* cancerous tissue), as well as between pluripotent and differentiated cells, occur outside of CGIs but within 2kb from their boundaries, in regions that were named CGI shores [29, 30]. Furthermore, cancer has been associated to a loss of sharply defined boundaries

between low- and high-methylated areas (CGI and CGI shores). In particular, methylation hypervariability in CGI shores, when found in phenotypically normal tissue, is predictive of tumorigenesis onset [31].

When considering the effects of a global (rather than a local) loss in DNA methylation, genomic instability has long been thought to derive from re-mobilization of normally silent transposon elements [32]. Also loss of imprinting (LOI) events, by relieving repression on entire chromosomes or part of chromosomes, could lead to overexpression of oncogenes and silencing of tumor suppressors [33]. Notably, these phenomena have also been attributed to loss of repressive histone marks such as tri-methylation of lysine 9 on histone H3 (H3K9me3), tri-methylation of lysine 20 on histone H4 (H4K20me3) and tri-methylation of lysine 27 on histone H3 (H3K27me3) [34-36].

Indeed, it is now clear that a crosstalk between DNA methylation and histone modifications exists (mediated at least in part by interactions between histone- and DNA-methyltransferases), suggesting an interdependence of the two types of epigenetic mechanisms [37]. For example, CGI promoter hypermethylation, associated to gene repression, often coincides with loss of histone marks typical of active genes (including H3 and H4 acetylation and H3 lysine trimethylation) [37]. A decade ago, Vogelstein and colleagues reported that tumor suppressor gene silencing appears to be induced by histone modifications prior to changes in DNA methylation [38]. The hypothesis of tumor suppressor silencing being primarily correlated with histone modifications rather than DNA methylation has been further supported by an increasing number of studies [37, 39-41]. This would be in agreement with reports showing that CGI hypermethylation in cancer occurs at genes that have already been silenced in non-tumoral tissue, and that it normally follows chromatin modification during development [40, 42, 43].

Surprisingly, out of >60 histone residues for which modifications have been identified, only a few have been associated to cancer to date, along with alterations in histone-modifying enzymes expression or activity [44]. The first histone mark deregulation found in cancer has been the combined loss of global acetylation of histone H4 at lysine 16 (H4K16Ac) and H4K20me3, occurring (along with DNA hypomethylation) at

repetitive sequences, and correlating with tumor progression [45]. H4K16Ac is thought to influence histone-chromatin interactions and to regulate chromatin folding, while H4K20me3 is usually linked to heterochromatic regions [46, 47].

Nevertheless, the most frequently altered modifications in tumors have been reported to be di- and tri-methylation of histone H3 at lysine 4 (H3K4me2/me3), normally positioned near transcription start sites and associated to transcriptional activation, and found to be decreased in a range of cancer types [48-51]. This finding is accompanied by the observation of an aberrant expression or activity of several enzymes controlling H3K4me2/me3 levels, such as the histone demethylases (HDMs) LSD1 and JARID1 and the histone methyltransferases (HMTs) of the MLL family (in the latter case due to protein fusion) [52-55]. Another mark commonly altered in cancer is H3K27me3, linked to repressive chromatin domains and sustained by the Polycomb repressive complex 2 (PRC2). In this case, both evidence of increase and decrease of activity of H3K27me3-controlling enzymes (such as the EZH2 component of PRC2, and the HDMs JMJD3 and UTX) and of levels of H3K27me3 have been reported, suggesting the importance of a proper equilibrium for this modification in the maintainance of normal cell growth [56-60]. The same bivalency has been described for H3K9me3, a typical mark of heterochromatin and transcriptional repression [61-64].

As in the case of DNA methylation, also disruption of normal histone modification patterns may have impacts on genome instability, and cause relevant changes to gene expression and higher order chromatin interactions. The balance between euchromatic and heterochromatic histone marks appear to be critical to this respect, and cancer therapies focused in repairing the loss of this equilibrium (such as inhibitors of HMTs, histone acetyltransferases or deacetylases) are being increasingly considered [65], though the precise rational bases for some of these therapies is still unclear.

1.2.1 The epigenetic cancer progenitor model

A wealth of epidemiological evidences has associated pre- and postnatal environmental factors to risk of adult development of several chronic diseases, including cardiovascular diseases, obesity, diabetes and cancer [66, 67]. Furthermore, cases of environmental prenatal and early postnatal environmental factors (such as

nutrition [68], xenobiotics [69], and low-dose radiation [70]) linked to aberrant epigenetic programming and higher disease risk have been frequently reported in literature. On the basis of these observations, Jirtle and Skinner proposed the "developmental origins of adult-onset disease" theory to account for the relationship between early environmental inputs and epigenetic alterations [71]. According to this model, the early adaptation of an organism to environmental influences (referred to as "developmental plasticity") can lead to a higher risk of chronic disease development in cases where the perceived environment will differ from the one experienced subsequently in adulthood.

In relation to tumor biology, the hypothesis of an "epigenetic progenitor" origin has been proposed [23], stemming from the observation of the reversibility of cancer phenotype in given conditions [72, 73]. Indeed, several studies demonstrated that tumor cells could revert to normal phenotypes upon exposure to specific environmental cues, a phenomenon that would be hardly explained in the context of permanent, genetic mutations [74, 75]. According to the epigenetic progenitor model, the first step of tumorigenesis would involve the epigenetic disruption of the progenitor cell, which would favor the occurrence of an initiating mutation and the acquisition of an enhanced genetic and epigenetic plasticity. In this light, reversible (epigenetic) changes could lead to irreversible (genetic) modifications, possibly by uncovering detrimental genetic variants (including transmissible ones) that are normally kept silenced by epigenetic mechanisms [23]. More recently, these concepts have been further extended [41], by proposing cancer as the consequence of an "epigenetic dysregulation". In this condition, the epigenome is endowed with a higher degree of plasticity, explaining the high phenotypic variability occurring among individual lesions, and favoring the natural selection of altered cells on a much shorter timescale than the one allowed by mutation variability [41].

The model of the "epigenetic progenitor" is gradually receiving confirmation from emerging evidences of an early involvement of epigenetic aberrations in tumorigenesis. In colorectal cancer, LOI of the insulin-like growth factor 2 (IGF2) mitogen is detected in both normal and neoplastic tissue, and induction of *Igf2* LOI increases tumor frequency [76]. Epigenetic alterations are identified at the earliest

steps of epithelial carcinogenesis in Barrett's oesophagus and in cervical cancer, largely before invasive malignancy onset [31, 77]. More recently, a large study found alterations in DNA methylation in the normal tissue of patients affected by different cancer types, with the number of alterations being directly correlated with age, providing a mechanism for the age-dependency of cancer frequency [78].

1.3 The interplay between genetic and epigenetic factors in cancer

The first case of genetic disruption of an epigenetic regulator in the context of cancer was identified in 1998 in malignant rhabdoid tumors caused by the mutation of the chromatin remodeler SMARCB1/SNF5 [79]. In the following years and especially after the diffusion of high-throughput sequencing studies, many other epigenetic modulators have been found mutated in cancer. Genes involved in epigenetic mechanisms represent half of the most significantly mutated genes in medulloblastoma [80], and are genetically altered in half of bladder cancers [81] and hepatocellular carcinomas [14]. The group of epigenetic modifiers most frequently targeted by mutation in common solid tumors is the chromatin remodeling class (including ARID1A, mutated in more than ten cancer types), whose inactivation causes an increase in euchromatic regions and in consequent gene activation [41, 82].

Nevertheless, most of the mutations on epigenetic modifiers isolated so far in solid tumors belong to either pediatric lesions (as in the case of childhood glioblastoma, having a 35.6% frequency of mutations on the histone variant H3F3A, in contrast to the 3.4% frequency in adult glioblastomas [83]), or rare, more aggressive variants of adult cancers (such as the rare pancreas neuroendocrine cancer, showing a 44% mutation frequency on the HMT multiple endocrine neoplasia 1 (MEN1), differently from the common pancreatic adenocarcinoma which only associates to an 8% frequency of mutation of the histone acetyltransferase p300 [84]). On the contrary, genetic disruption of epigenetic players is a frequent event in hematological cancers, for example in acute myeloid leukemia and lymphoma [85-87].

These observations seem to indicate that, in common solid tumors, epigenetic abnormalities are a more frequent tumorigenic mechanism than genetic mutations in

epigenetic regulators. The overrepresentation of epigenetic-associated mutations in highly aggressive and rare types of lesions further suggests genetic derangement of epigenetic machineries to have a particularly heavy impact on cellular activities [41].

Epigenetic mechanisms, in turn, can play a fundamental role in the occurrence of malignant genetic modifications, to the point of influencing a tumor's mutation rate. Several examples of such condition have been reported in literature [88-95].

Apart for relieving the repression of transposon sequences in the genome, DNA hypomethylation has been associated to higher frequencies of structural mutability, and hypomethylated regions have been found enriched in DNA breakpoints leading to copy number variations [88-90].

DNA methylation has been found to have a protective effect against DNA double strand breaks induced by guanine quadruplexes (G4s), four-stranded structures deriving from folding of G-rich sequences ($G_3N_{1-7}G_3N_{1-7}G_3N_{1-7}G_3$). Hypomethylated regions frequently include DNA breakpoints, and these are often associated to an enrichment in G4s [96].

DNA methylation also has an impact on methylation-dependent mutations deriving from spontaneous hydrolytic deamination. In contrast to methyl-cytosine deamination, which yields a thymine, unmethylated cytosines produce uracils (easily recognized as abnormal bases in DNA, and therefore repaired). In highly proliferative tissues, deamination of methyl-cytosines in parent strands just before DNA replication (when the two DNA strands are detatched) prevents recognition of the lesion and give rise to a full C:T substitution. Consequently, the rate of C:T transitions on CpG dinucleotides, which are frequently methylated, is ~10 times higher than any other single nucleotide variant (SNV) in human genome [91], and 25% of all mutations on TP53 in cancer have been attributed to this epigenetic modification [92].

Chromatin modifications also correlate with mutation rates. High 3K9me3 levels, typical of in large repressive domains, are associated to >40% of SNVs in human cancers [93], although at least partially due to a lower evolutionary constraint proper of heterochromatin [91]. Common fragile sites (CFS), large unstable genomic regions sensitive to replication perturbations and variations in copy number, have been found to be hypoacetylated [94]. Finally, genome instability can also originate from

epigenetic silencing of components of the DNA repair machineries, such as MLH1, MGMT, BRCA1, WRN, FANCF, and CHFR [95].

Clearly, the genome and epigenome impact on each other in a variety of ways, providing complementary mechanisms to reach comparable effects in cancer.

Several reports described that mutational inactivation and epigenetic silencing of the same gene are two mutually exclusive phenomena, in analogy to the occurrence of two mutations on the same pathway. Thus VHL, a key tumor suppressor for clear cell renal carcinoma, was found mutated in 60% of tumors and silenced by hypermethylation in another ~20% [97], and mutation and epigenetic repression are mutually exclusive also in the case of E-cadherin in the lobular and ductal forms of breast cancer, BRCA1 in ovarian cancer, and CDKN2A in squamous cell lung cancer [98-100].

Nevertheless, genomic and epigenomic changes can act cooperatively, as in the case of the CpG island methylator phenotype (CIMP) and BRAF mutation in colon cancer, where CIMP, consisting in high frequencies of DNA hypermethylation on a specific set of CGIs, generates a favorable context for BRAF mutation at the early precursor stage [14, 101].

Altogether, these evidences point to a strong correlation between areas of genetic aberration and epigenetic derangement, implying a collaborative effort of mutations and epigenetic modifications along the tumorigenic process.

2. Inflammation

Inflammation is an immediate adaptive attempt to restore tissue homeostasis upon noxious insults, including pathogenic infection and physical injury. Acute inflammation represents an immediate and short-term type of response, whereby leukocytes infiltrate the affected site, remove the triggering stimulus and repair the damaged tissue. This well-characterized, physiological response might become pathological if not properly terminated. Indeed, chronic inflammation is a persistent, maladaptive response composed of active inflammation, tissue destruction and concurrent repair. This condition is much less understood, and nevertheless lies at the basis of a number

of chronic human disorders, such as allergy, atherosclerosis, autoimmune diseases, and cancer.

2.1 Inflammation is a cancer hallmark

Chronic inflammation has been causally associated to cancer as early as in the 19th century by Rudolf Virchow, upon observation of the presence of leukocytes within tumors. In time this association received growing support from the literature, to the point that tumors have been defined as "wounds that do not heal" [102]. A plethora of epidemiological evidences linked chronic inflammation to different forms of cancer. Up to 20% of tumors are correlated to microbial infections (for example, chronic viral hepatitis and hepatocellular carcinoma, *H. pilori* and gastric cancer, HPV and cervical cancer), 35% can be ascribed to nutrition factors (20% being associated to obesity, now acknowledged to be consistently linked to inflammation [103]), 30% can be related to inhaled chemicals triggering inflammation (such as smoke, asbestos and silica), and many others to autoimmune disorders (as for the case of inflammatory bowel disease and colorectal cancer) [104, 105]. Furthermore, inflammatory components (cells, cytokines and chemokines) infiltrate most tumor lesions, including those not directly caused by an underlying inflammatory condition [106]. Anti-inflammatory drugs have been reported to prevent tumor onset or delay tumor progression in specific cases, notably colon cancer [107].

Many interesting insights have been uncovered on the molecular events linking chronic inflammation to tumor formation and progression. The emerging picture suggests a feed-forward loop where activated innate immune cells stimulate tumor growth and progression via secretion of cytokines and pro-inflammatory mediators, and cancer cells produce soluble mediators that recruit and activate inflammatory cells, thus further fostering tumor progression [108].

Furthermore, pro-tumoral inflammation and anti-tumoral immunity coexist at all the different steps of the tumorigenic process, and environmental and microenvironmental cues control the balance between them [109]. The composition of the cytokine and chemokine pool secreted in the inflammatory milieu is particularly important in governing this delicate equilibrium [109]. By activating a variety of downstream

effectors, including NFκB (nuclear factor kappa-light-chain-enhancer of activated B cells), AP1 (activating protein 1), STAT (signal transducer and activator of transcription) and SMAD (mothers against decapentaplegic homolog) family transcription factors, cytokines can shift this balance to either promote anti-tumoral immunity (interleukin (IL) 12, IL1α, interferon (IFN) γ, tumor growth factor (TGF) β), or stimulate cancer progression (tumor necrosis factor (TNF) α, IL6, IL1β), although clear-cut separation between pro- and anti-tumoral players is not always possible (**Figure 1**). The bioactive molecules that typically compose an inflammatory microenvironment promote tumorigenesis at multiple levels, with growth factors supporting mitogenic signaling, survival factors inhibiting apoptosis, angiogenic factors stimulating neovascularization, enzymes that can remodel the extracellular matrix helping invasiveness and metastasis [110, 111]. Indeed, the idea that cancer biology could not be exhaustively comprehended considering cancer cells only, but needs to entail also the tumor microenvironment, allowed for inflammation to be acknowledged as one of the cancer hallmarks itself [112].

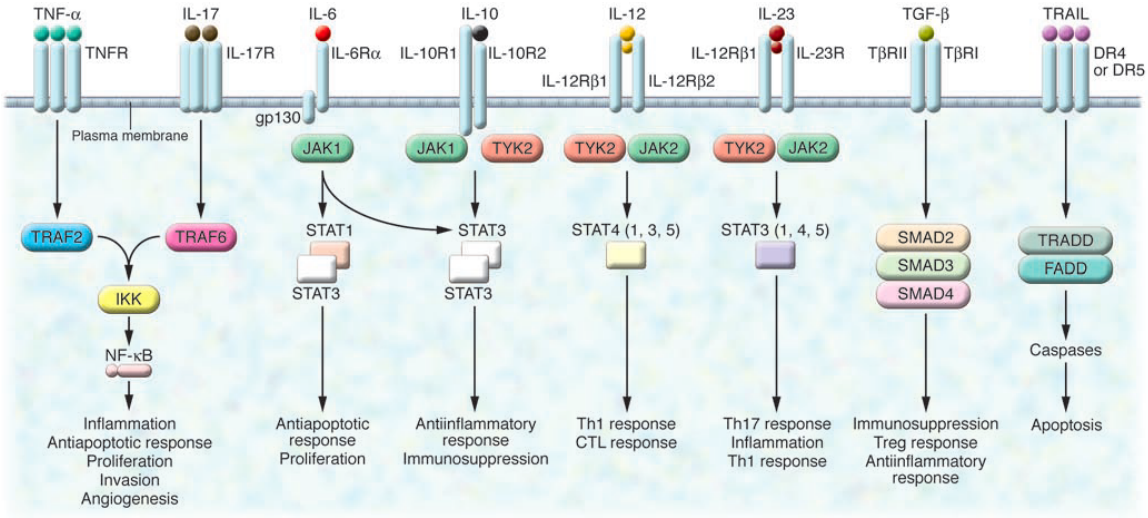


Figure 1

Major signaling pathways and downstream responses of inflammation-related cytokines in cancer. DR4 = death receptor 4, FADD = Fas-associated death domain, gp130 = glycoprotein 130, TRAIL = TNF-related apoptosis-inducing ligand, TRADD = TNF receptor–associated death domain protein, TRAF2 = TNF receptor–associated factor 2, TYK2 = tyrosine kinase 2. Adapted from [109].

2.2 Inflammation and the acquisition of genetic changes in cancer

Inflammation is frequently observed at the earliest steps of a tumorigenic process, and its ability to foster progression of an early lesion into a frank tumor is out of discussion. Nevertheless, a precise assessment of the impact of an inflammatory response on tumor onset is challenging. This is partly due to the lack of proper *in vivo* models enabling a clear evaluation of possible effects on the earliest cancerous stages. Furthermore, our current knowledge mostly relies on observations performed starting from a detectable tumor load, when probably escape of immunitary surveillance had already taken place.

Several lines of evidence suggest that inflammation is associated to genome destabilization. In a rodent model of bowel inflammation deriving from ablation of the anti-inflammatory IL10 cytokine, mutation frequencies in the colon were increased by 4-5 fold [113]. A higher mutation rate has been observed also for inflamed tissues in non-tumoral conditions. In a comparison of human tumors and matched healthy tissues, the only exception to the much lower mutation frequency scored by normal tissues was an inflamed sample, reaching a rate of $4*10^{-8}$ mutations/bp (in contrast to $<1*10^{-8}$ mutations/bp for the non-inflamed samples) [114]. Microsatellite instability and high frequencies of p53 mutations have been reported in the tissues of non-oncologic patients affected by pancreatitis, ulcerative colitis, rheumatoid arthritis and atherosclerosis [115-117]. Copy number variations, LOI and mutations on tumor suppressor genes have been also documented in tumor-related stroma, although these findings are controversial [118-120].

Yet, in spite of these correlative data, a clear proof of a direct mutagenic potential of inflammation is still missing. Activated macrophages and neutrophils produce ROS and RNS that can cause DNA damage, and that are therefore obvious inflammation-associated candidates to explain genomic instability. Nevertheless, whether ROS and RNS released by inflammatory cells are stable enough to react with DNA packed into chromatin after having diffused through the extracellular matrix, penetrated a cell, crossed its cytoplasm, and entered the nucleus, is still unclear. A possibility is that immunity cells could induce ROS accumulation in the surrounding epithelial cells via TNFα secretion. However, the activation of scavenging mechanisms during chronic

inflammation should dampen the ability of ROS and RNS to cause damage to cellular structures.

Besides causing the production of genotoxic compounds, inflammation could also drive mutagenesis by impacting on DNA repair systems. Members of the mismatch repair machinery are downregulated by ROS and RNS, which is associated to an increase in microsatellite instability [121], and few evidences suggest an impact also on the base excision and the nucleotide excision patways [121].

Another mechanism that could provide a rationale for an inflammatory mutagenic property is linked to AID, a DNA/RNA editing enzyme physiologically responsible for the immunoglobulin gene class hypermutation in B cells. AID has been found overexpressed in a number of tumors and surrounding inflamed tissues, such as chronic gastritis and gastric cancer, inflamed colon and colitis-associated cancer, inflamed liver and hepatocellular carcinoma [122-124]. AID expression is dependent on several cytokines, including TNFα, IL1β and TGFβ, and has been described to induce aberrant double strand breaks (and consequently, mutations and translocations) on Myc and BCL6 [125, 126].

Finally, an intriguing possibility is that inflammation could lead to the production of cytokines and growth factors that would impart stem cell-like properties to tumor progenitors and foster the expansion of the stem cell pool. In this regard, a recent work has shown that ROS, apart from acting as damaging agents, can also act as signaling molecules and induce NFκB-dependent initiation of colon cancer by intestinal stem cells at the crypts base [127].

However, final proof that any of these inflammation-driven phenomena actually make a crucial contribution to cancer initiation is still missing, and further studies will be necessary to unequivocally determine which mechanism is at the basis of inflammation mutagenicity.

2.3 Inflammation and epigenetics

Since aberrant epigenomic changes have been causally associated to oncogenesis, an impact of chronic inflammation on epigenetic mechanisms has been hypothesized as a causing event for inflammation-driven tumor initiation. This connection would also

provide a further explanation to how environmental cues could trigger a tumorigenic process, as postulated by the epigenetic progenitor model [23]. In particular, the achievement of a more dynamic epigenetic profile (that would eventually allow for the acquisition of a malignant epigenetic pattern) has been proposed to derive from the reiterated rearrangement of the epigenetic landscape occurring in an inflammatory context [41].

Both inflammation and alterations in epigenetic control are early features of a malignant lesion, and are often reported to take place at pre-neoplastic stages. Yet, whether they act concomitantly to promote cancer initiation, or if one of the two processes is secondary and dependent on the former, it is still not completely clear. Nevertheless, evidences in support of epigenetic consequences of an inflammatory condition are now starting to accumulate.

ROS and RNS have been found implicated also in modulating epigenetic regulators. For example, oxidative stress has been described to cause the activation of silenced genes by inactivation of histone deacetylases (HDACs) [128]. It can also favor the oxidation of 5-methylcytosine to 5-hydroxymethylcytosine, which is not recognized by the DNA methyltransferase 1 (DNMT1) and therefore leads to methylation loss after mitosis [129]. On the contrary, halogenated byproducts of inflammatory responses could result in a methylation gain. Specifically, HOCl (produced by neutrophils myeloperoxidase) and HOBr (deriving from eosinophils peroxidases) interact with unmethylated cytosines yielding 5-chlorocytosines and 5-bromocytosines, respectively. These two halogenated cytosines mimick the 5-methylcytosine and cannot be distinguished by DNMT1, which therefore inserts a stable methylation at the modified site [129].

DNA methylation levels have been shown to undergo alterations during inflammation also independently of oxidative stress phenomena. A global hypermethylation has been observed in non-cancerous tissues of oncologic patients with respect to healthy individuals [130], and during systemic inflammation in patients affected by chronic kidney disease or atherosclerosis [131, 132]. Furthermore, several reports show IL6 to modulate DNA methylation in erythroleukemia, cholangiocarcinoma and oral cancer [133-135].

Hahn and colleagues found ~60% of cancer-associated DNA methylation changes to be present already in pre-malignant, inflamed tissue, and that the degree of inflammation-specific DNA methylation retained in tumor is much more significant than the age-dependent DNA methylation [136]. The same study showed that 70% of the genes associated with inflammation-dependent DNA methylation are Polycomb targets, and that acquisition of DNA methylation at CGI seems to be directed by Polycomb and to correlate frequently with a concomitant loss in H3K27me3 [136]. In addition, Jmjd3, an H3K27me3-specific HDM, was demonstrated to be induced by inflammatory stimuli in macrophages [137]. More recently, induction of colitis in a mouse model of colon cancer changed (most often decreased) H3K27me3 levels at ~3600 regions, with some of these changes being retained along the progression to tumor [138].

All these data support the hypothesis that chronic inflammation may promote a progressive deterioration of epigenetic marks, in turn favoring cancer development. Further studies will help to better characterize the interplay and the temporal relationship between inflammation and epigenetic derangement.

3. Liver cancer

The liver is the largest gland in the body, accounting for 2%-5% of body weight. It exerts a plethora of essential functions, such as production of endocrine (insulin-like growth factors, thrombopoietin, angiotensinogen), exocrine (bile) and plasma protein (albumin, apolipoproteins) secretions, storage of glycogen, modulation of cholesterol synthesis and transport, drug detoxification and control of metabolism. Given its crucial role on the maintenance of physiological homeostasis, liver pathologic conditions (including fibrosis, cirrhosis, hepatitis, cholestasis, and hepatocarcinoma) are generally associated to high morbidity and mortality rates. In particular, liver cancer is the 6th commonest malignant lesion and the 3rd cause of cancer-related death worldwide [139], hepatocellular carcinoma (HCC) being the principal histological type. Differently from other carcinomas, an increase in its incidence is observed, with ~750,000 new cases and ~700,000 deaths reported globally every year. Survival

rates did not significantly rise over the last three decades, and the available therapeutic approaches are yet currently limited to surgery, percutaneous and transarterial interventions, and treatment with sorafenib, a multikinase inhibitor blocking the MAPK pathway [140].

The lack of targeted molecular therapies is largely due to the complexity and heterogeneity of HCC, and the presence in a large fraction of cases of concurrent liver disease. Indeed, >80% of HCC lesions originate in a context of liver cirrhosis [141]. Other pathological situations commonly associated to HCC include nonalcoholic steatohepatitis, nonalcoholic fatty liver disease, autoimmune hepatitis, diabetes mellitus, hereditary metabolic liver diseases (e.g. α1-antitripsyn deficiency, hemochromatosis, glycogen storage disease type I) and obesity [142-147]. Nevertheless, the vast majority of HCCs develop after chronic infection from HBV, HCV or HDV or chronic exposure to toxins (e.g. alcohol, aflatoxins) [148].

3.1 Human HCC has been sequenced

The complexity of liver cancer mutational repertoire has been underlined in a recent study, where it has been associated to the maximum number of mutational signatures identified for a cancer class: whereas most tumor types display two possible mutational signatures, liver tumor reaches six signatures [16]. Indeed, many genes are altered in HCC, but the frequency of individual gene mutations is low, therefore no universal genetic modification has been found in HCC. Nevertheless, the application of NGS technologies disclosed previously uncharacterized mutation patterns and chromosomal alterations, shedding light on the genetic heterogeneity deriving from different HCC etiologies.

A number of recent publications report whole-genome sequencing (WGS) or whole-exome sequencing (WES) data on HCC [14, 149-155], in patients with either viral infection (HCV or HBV) or chronic exposure to alcohol.

Mutation frequencies stated in these articles are in line with other solid tumors. Up to 3739 genes have been found mutated in the analyzed patients [155], and several genes were recurrently mutated (*Tp53, Ctnbb1, Arid1a* and *2, Axin1, Mll, Tert*), most of them being well-known oncogenes [14, 149-155].

The major types of base substitution observed in these samples are T>C/A>G transitions and G>T/C>A transversions [14, 150, 152, 154]. Interestingly, G>T/C>A transversions are known as a hallmark of aflatoxin B1 mutagenic exposure [156]. Furthermore, Guichard and colleagues observed that G>T/C>A substitutions were significantly more frequent in lesions stemming in a non-cirrhotic liver [152], and Huang et al. reported them to be common in HBV-derived HCCs with portal vein thrombosis [154]. On the other hand, T>C/A>G transitions at ApT sites have been primarily found in HCV-related HCCs [150]. Additionally, also chronic alcohol consumption and occurrence of multiple liver nodules displayed a significant association with components of the somatic substitution pattern [14].

These genomic studies also generally reported a high number of small indels (as many as 670 in the work by Totoki and colleagues [150]).

Frequent copy number variations (CNVs) have been detected [14, 149, 152, 153, 155], often spanning entire chromosome arms, and with copy number deletions being generally predominant over amplifications. Several rearrangements (including gene fusions) have been listed, mostly at copy number boundaries [14, 150, 155]. Interestingly, Sung *et al.* noted that CNVs were mostly accumulating at HBV integration breakpoints [153], and Fujimoto *et al.* found HBV to integrate most often within or upstream of TERT [14], thereby highlighting the contribution of viral integration events as a source of chromosomal instability and as a specific cancer signature.

The top 5 pathways recurrently altered were stated to include, in order of frequency, the Wnt/beta-catenin, p53/apoptosis and cell cycle control, chromatin remodeling, PI3K/Ras and oxidative and endoplasmic reticulum stress pathways [152]. In particular, mutations in ≥1 gene involved in the chromatin-remodeling pathway, including ARID1A, ARID1B, ARID2, MLL, MLL3, BAZ2B, BRD8, BPTF, BRE, HIST1H4B and SMARCA genes, were found in 24% [152] and 52% [14] of the HCC cohorts taken into account.

CTNNB1 mutations were found to be more frequent in HCV- than in HBV-driven HCCs, and to be mutually exclusive with TP53 mutation [151, 152]. Also mutations on the gene encoding for ARID2, a subunit of a SWI/SNF chromatin-remodeling complex,

were estimated to occur more frequently in HCV-HCCs and to be mutually exclusive with mutations on TP53, and it has been reported to be correlated with CTNNB1 mutations [151]. Nevertheless, Guichard and colleagues observed that mutations on ARID1A, another member of the ARID protein superfamily, despite being associated to CTNNB1 mutations and mutually exclusive with ARID2 mutations, were more frequently detected in alcohol-related HCCs [152].

On the contrary, mutations leading to inactivation of IRF2, and consequent disruption of TP53 function, were specifically associated to HBV-derived lesions [152]. Moreover, a higher degree of CNVs and chromosomal instability has been reported in HBV-related lesions, HCCs having TP53 mutated and tumors developed in a non-cirrhotic background [152].

Finally, mutations on AXIN1 and APC genes were observed in HCCs with different etiologies. However, mutations on RPS6KA3, a ribosomal protein kinase involved in PI3K/Ras signaling, were associated to AXIN1 mutations and found mainly in tumors without cirrhosis [152].

All together, these findings shed light on the impact of etiology on the mutagenic pattern of each individual HCC lesion. Nevertheless, most of the tumors analyzed in these studies were originating from an environmental cause, including viral infection and exposure to chemicals. Such a pervasive impact of external agents on the genomics alterations acquired in HCC has so far prevented a definitive understanding of the direct contribution of conditions such as cirrhosis, fibrosis, and, in general, chronic inflammation to disease initiation and progression. Therefore, despite the clear inflammatory component of the tumorigenic process triggered by these factors, the impact of inflammation *per se*, independently of other associated etiologic elements has yet to be investigated.

3.2 *Mdr2*-KO mice: a model of spontaneous, purely inflammatory tumorigenesis

The multidrug resistance 2 (*Mdr2*) knockout mouse (FVB.129P2-Abcb4^{tm1Bor}) represents a well-characterized model of spontaneous hepatocarcinogenesis in the context of chronic inflammation. *Mdr2* (or ATP-binding cassette B4, *Abcb4*), ortholog

of the human *Mdr3*, encodes for the P-glycoprotein responsible for flipping the phosphatidylcholine (PC) phospholipid from the inner to the outer leaflet of the hepatocytes canalicular membrane (**Figure 2**) [157, 158]. It is expressed at low levels also in spleen, skeletal muscle, adrenal gland, tonsil and heart, but no morphologic or histologic abnormality has been scored in districts other than liver in KO mice [157].

Under physiologic conditions, biliary PC is transported into bile ducts *via* MDR2 and subsequently forms mixed phospholipid-cholesterol-bile acid micelles, thus protecting cholangiocytes from injury caused by monomeric bile acid. The lack of biliary PC in *Mdr2*-KO mice therefore results in persistent biliary epithelium damage, ultimately leading to leaky bile ducts with regurgitation of bile acid into portal interstitium and consequent inflammatory response. Furthermore, the stability of mixed micelles depends on a proper proportion of phospholipids and bile salts, required to maintain cholesterol solubility. Destabilization of micelles due to absence of PC would favor crystallization of cholesterol and bile lithogenicity, leading to obstruction of bile canaliculi and consequent ductural proliferation [159].

The pro-inflammatory microenvironment originating in the context of chronic portal inflammation impacts on the periportal hepatocytes causing various degrees of injury and apoptosis, with subsequent regenerative hyperplasia. Accelerated DNA replication is *per se* error prone and facilitates the occurrence of further pro-tumorigenic alterations. On the other hand, pro-inflammatory mediators (particularly TNFα, CCR5 and NF-κB) derived from the persistent portal inflammation also play a pivotal role during the malignant transformation of proliferating hepatocytes and consequent tumor progression [158, 160, 161].

Similarly to human hepatocellular carcinoma, tumor development in *Mdr2*-KO mice progresses through defined phases, starting from a chronic inflammatory condition that ensues at an early stage (as soon as 8 weeks of age) and is characterized by periductular mixed inflammatory infiltrate (rich in CD3+ cells), bile duct proliferation and hepatomegaly [160, 162]. When mice are approximately 4 months old they present portal tract expansion caused by bile duct hyperplasia, and start to be affected by severe architectural and cytologic liver dysplasia. At 10 months, multiple

adenomatous lesions appear, which later (between 12 and 16 months) give rise to grossly visible nodular carcinomas, with a 100% penetrance in the FVB genetic background [160, 162]. Since HCCs develops within adenomas [163], each neoplastic nodule offers a screenshot of the adenoma-to-carcinoma transition and displays a variable HCC fraction (histopathologically measured as "tumor content"), which tends to increase with time.

The existence of clearly identifiable pathological phases, along with the independence from external mutagenic stimuli, makes the *Mdr2*-KO mice a suitable model for the study of the impact of inflammation on genetic and epigenetic variations occurring over time during a tumorigenic process.

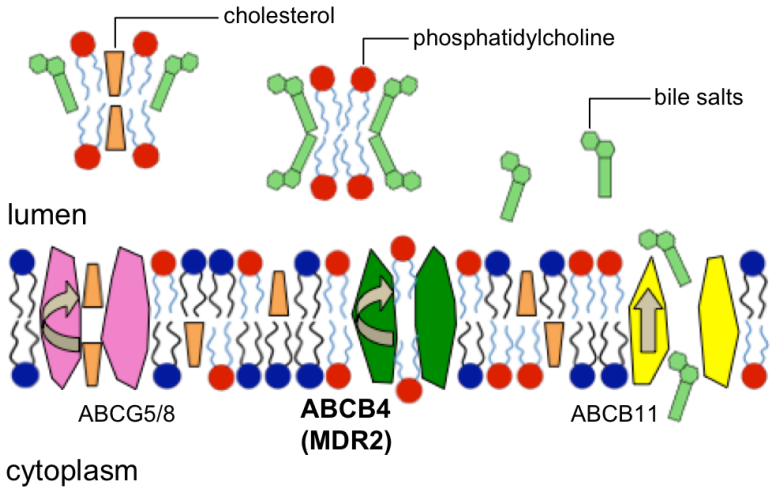


Figure 2

Function of the murine MDR2 floppase on the hepatocyte apical membrane. Adapted from [164].

3.3 Progressive familial intrahepatic cholestasis (PFIC)

The term "progressive familial intrahepatic cholestasis" (PFIC) was originally used to discriminate cholestatic end stage liver disorders in children that could not be ascribed to neonatal or pediatric cholestasis such as biliary atresia, Aagenaes syndrome or Alagille syndrome. It now defines a heterogeneous class of rare autosomal recessive liver diseases, whose common feature is represented by neonatal hepatic cholestasis that progresses to fatal liver failure occurring during childhood or adolescence. Although rare, PFIC is associated to 10-15% of pediatric cholestasis cases and of liver

transplantation in children. PFIC has an estimated incidence of 1/50,000 to 1/100,000 births, with no geographical or gender preference [165].

		PFIC 1	PFIC2	PFIC3
Gene		ATP8B1	ABCB11	ABCB4
Protein		FIC1	BSEP, SPGP	MDR3
Chromosomal locus		18q21-22	2q24	7q21
Function		PS-flippase/microvillus formation	Bile salt export	PC-floppase
Cholestatic onset		1st year - adolescence	1st year - adolescence	1st year - early childhood
Biliary bile acid		Very low, PS present	Absent	Normal but devoid of PC
Serum	**Bile salts**	High	High	Normal
	γGT	Normal	Normal	High
	ALT	Moderately increased	High	High
	Cholesterol	Occasionally high	Recurrently high	Normal
Liver histology		Canalicular cholestasis, granular bile	Giant cell hepatitis	Bile duct proliferation and fibrosis
Pruritus		Severe	Severe	Moderate
Extrahepatic symptoms		Diarrhoea, pancreatitis, hearing loss	None	None

Table 1

Description and distinguishing features of the three types of PFIC. Adapted from [166] and [165].

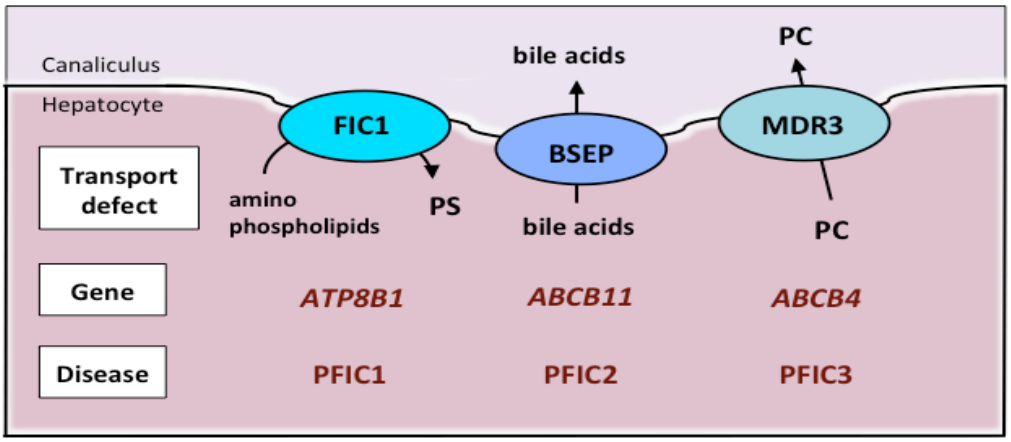


Figure 3

Roles of the hepatocyte transporters mutated in PFIC. PS = phosphatidylserine; PC = phosphatidylcholine.

Molecular studies allowed to distinguish 3 forms of PFIC, based on the originating genetic lesion. In each case, mutation of a gene encoding for a hepatocellular transporter is involved (**Table 1**, **Figure 3**).

3.3.1 PFIC1

PFIC1 is due to mutation in the *Atp8B1* gene (also called *Fic1*, familial intrahepatic cholestasis 1), which encodes for a P-type ATPase located on the canalicular surface of hepatocytes, but also on the apical membrane of cholangiocytes and in organs such as intestine and pancreas.

ATP8B1 has been recently demonstrated to act as a flippase by transporting a membrane phospholipid, phosphatidylserine (PS), from the outer to the inner membrane leaflet, where PS is generally confined [167]. ATP8B1 activity is probably important to control membrane fluidity and to confer resistance to bile salts. Indeed, it prevents PS from diluting, along with the excess PC flopped by ABCB4, the concentrations on the outer leaflet of sphingolipids and cholesterol, thereby allowing the leaflet to form a liquid crystalline phase that would be less sensitive to detergent solubilization.

PFIC1 histological features consist in a generally preserved lobular architecture, and are therefore associated to a milder damage with respect to the ones observed for PFIC2 and 3 [168]. Accordingly, serum levels of gamma-glutamyltranspeptidase (γ-GT) and alanine aminotransferase (ALT), enzymes routinely screened as markers of hepatic damage, are only slightly elevated in PFIC1 patients. The most remarkable finding for PFIC1 is the so-called "Byler bile", coarse granules of bile accumulating into hepatocytes [168]. Although evident, signs of cholestasis are limited to the peri-canalicular areas. Yet, PFIC1 patients typically develop cirrhosis and severe liver failure by their second decade.

The mechanism linking ATP8B1 mutations to cholestasis consists in a significant downregulation of farnesoid X receptor (FXR), a nuclear receptor regulating bile metabolism, which determines a downregulation of BSEP, the bile acids transporter in the liver, and a parallel upregulation of bile acid synthesis as well as of the apical sodium bile salt transporter (ASBT) involved in the bile acids reabsorption in the intestine [169]. All together, these events cause bile acid overload in the hepatocytes, which is in agreement with the low biliary bile acid concentration measured in patients [170].

Atp8B1 gene expression occurs in various districts other than liver, for example in

enterocytes (at the level of ileum and jejunum), cochlear hair cells in the ear and pancreatic acinar cells, and its ablation causes symptoms such as severe diarrhoea, hearing loss and pancreatitis, respectively. The fact that *Atp8B1* expression is higher in intestine than in liver [171] suggests it to be also involved in the bile salts enterohepatic cycling. This would explain the chronic diarrhea in a number of PFIC1 patients.

3.3.2 PFIC2

The second form of PFIC is due to mutations in the *Abcb11* gene, expressing the ATP-dependent canalicular bile salt export pump (BSEP, also called sister of P-glycoprotein, SPGP) in the hepatocyte canalicular membranes. BSEP is the main exporter of primary bile acids, and works against considerable concentration gradients. In particular, it principally secretes monovalent bile acid species, including primary bile acids such as cholic acid (CA) and chenodeoxycholic acid (CDCA) and the secondary bile acid deoxycholic acid (DCA). Additionally, it transports ursodeoxycholic acid (UDCA) into bile [172]. Collectively, bile salt excretion is necessary for the removal of cholesterol, which is water-insoluble: 12-20 g of secreted bile acids are sufficient to eliminate 500 mg of cholesterol [173]. It is also required for the absorption of other hydrophobic substances such as vitamins D, E, K and A, and for the excretion of bilirubin, a byproduct of erythrocytes degradation.

Mutations on *Abcb11* therefore impair biliary bile salt secretion, which causes reduced bile flow and chronic bile salt accumulation in hepatocytes, progressively leading to hepatic damage. Despite linked to extremely low biliary bile salts, PFIC2 is associated to normal serum γ-GT and cholesterol levels. Disease typically becomes evident within 6 months of life, with symptoms including jaundice, growth failure, pruritus, and occasionally vitamin K deficiency with consequent severe haemorrhage [165]. The most characteristic histologic finding in the early disease is the presence of giant cell transformation and different degrees of lobular and periportal inflammation [174]. Of note, PFIC2 is the only type of PFIC that has been reported to degenerate into hepatocellular carcinoma or cholangiocarcinoma [175-177].

3.3.3 PFIC3

PFIC3 is caused by genetic defects in the *Abcb4* gene (also called *Mdr3*, the orthologous of the gene abrogated in *Mdr2*-KO mice). As in mouse, MDR3 P-glycoprotein is a phospholipid floppase involved in biliary PC secretion and is mainly expressed in the hepatocyte apical membrane. Consequently, the mechanism of cholestatic liver disease in PFIC3 patients is dependent on the absence of biliary phospholipids. Human *Mdr3* and murine *Mdr2* genes share more than 90% homology [178], and indeed, when knocked-in to *Mdr2*-KO mice, human *Mdr3* completely rescued PC excretion into the bile [179].

In contrast to PFIC1 and 2, PFIC3 patients display symptoms of cholestasis within the first year only in 1/3 of cases, and might rather become affected later in childhood or adolescence. Also, differently from PFIC1 and 2, PFIC3 is characterized by high γGT levels and total absence of PC from bile. Symptoms include jaundice, mild pruritus, hepato- and splenomegaly, failure to thrive and gastrointestinal bleeding caused by portal hypertension [180].

Histologic examination indicates ductular proliferation, portal fibrosis and mixed inflammatory infiltrate. Occasionally, lobular cholestasis and slight giant cell transformation can be scored. As the disease progresses, severe portal fibrosis and biliary cirrhosis are observed, with interlobular bile ducts found in most portal tracts. Cirrhosis eventually progresses to end-stage liver disease, and 50% of patients require liver transplantation at 7.5 years [180]. Nevertheless, neither periductal fibrosis nor biliary epithelium injury are observed, and no liver tumor has yet been reported in association with PFIC3 [159, 165].

3.3.4 PFIC-HCC: a human counterpart for *Mdr2*-KO HCC

As previously mentioned, PFIC2 is the only type of PFIC that has been associated to hepatocarcinoma development to date. In particular, PFIC2 patients have an increased risk of developing HCC at age of two years equal to 5-10% [177].

Interestingly, the cholestatic phenotype displayed by *Abcb11*-KO mice is less severe than the one described in PFIC2 patients, and is not generally associated to HCC development [181]. In particular, these animals display liver injury with high mortality

rate only upon feeding with a cholate-containing diet [182]. This is attributed to the more hydrophilic bile salts pool found in mice, which includes muricholic acid and atypical species such as 24-tetrahydroxy bile acids (not found in humans) [182]. Nevertheless, despite the striking reduction in the cholic acid export observed in these mice, the total output of bile salts is almost unaffected due to the parallel upregulation of *Abcb1a*, a different ABC transporter that seems to compensate for *Abcb11* absence in mice, but not in PFIC2 patients [183, 184].

PFIC-HCC tumors are generally characterized by increased levels of α-fetoprotein (AFP) and nuclear accumulation of p53 protein, but no nuclear accumulation of β-catenin [177]. Recently, development of multiple, well-differentiated dysplastic nodules in a PFIC2 patient has been reported [175], suggesting PFIC hepatocarcinogenic process to follow a discrete progression through dysplasia, adenoma and HCC similarly to what described for *Mdr2*-KO mice.

PFIC-HCC commonly stems from an early histologic pattern including lobular/portal inflammation, along with giant cell transformation and hepatocellular swelling and necrosis [174]. Also this hepatitis-like condition closely resembles the persistent portal inflammatory reaction observed at the early stages of *Mdr2*-KO disease.

Therefore, despite differences in some of the histopathologic features acquired by the murine and the human disease, HCC lesions developed by *Mdr2*-KO mice and PFIC2 patients share a very close etiologic background, deriving from a chronic inflammatory setting that gives way to a step-wise tumorigenic process.

4. Aims of the project

Tumor formation and progression are accompanied by genetic, epigenetic and transcriptional alterations, some of which are directly and critically involved in bringing about the abnormal properties of tumor cells. Nevertheless, how these different layers of genome organization and function evolve in emerging and established tumors, influence each other, and impact on cellular physiology and tumor progression, is essentially unknown. Lack of such integrated knowledge is a major hurdle to a definitive understanding of the molecular bases of cancer development, and thus to

the development of mechanism-based therapies.

Indeed, most of the systems biology-based models of cancer produced so far rely on a single level of complexity (e.g. transcriptome) [185, 186]. Furthermore, they take into consideration only one step (the terminal stage) of tumorigenesis, where the essential driving events are usually outnumbered by secondary (bystander) genetic and epigenetic alterations.

This project aimed at dissecting the events occurring at each temporal step of a tumorigenic process at the level of genetic and epigenetic modifications, in order to obtain information that will be necessary to determine their temporal relationship and interplay.

To this respect, we found particularly interesting to characterize the genetic and epigenetic phenomena leading to cancer formation from a chronic inflammatory condition. Indeed, as both inflammation and epigenetic alterations are acknowledged as early events of a tumorigenic process, and since the mechanistic link between inflammation and tumor onset is still unclear, we wanted to understand whether inflammation could lead to cancer development by causing a progressive disruption of epigenetic patterns, and how would this impact over the genome.

For these purposes, we chose the liver as a model organ for several reasons: (1) the relative homogeneity of the tissue (in which over 70% of the cells are hepatocytes); (2) the large amount of material available, that would allow for multiple parallel experimental analyses; (3) the availability of several mouse models of inflammation-driven liver cancer. To this respect, the existence of clearly identifiable pathological phases, along with the independence from external mutagenic stimuli, makes the *Mdr2*-knockout mice a suitable model for the study of the genetic and epigenetic variations occurring over time during a tumorigenic process.

MATERIALS AND METHODS

Murine liver tissues and primary hepatocytes preparation

Founders of the FVB.129P2-Abcb4^{tm1Bor} (*Mdr2*-KO) and FVB.129P2-Abcb4^{wt} (*Mdr2*-WT) mice were purchased from The Jackson Laboratory. During initial pathological stages (inflammation, dysplasia), or on *Mdr2*-WT animals, downstream manipulations have been carried out on purified populations of hepatocytes obtained *via* collagenase liver perfusion, using a two-step protocol adapted from the original procedure proposed by Berry and Friend [187] and subsequently modified by Seglen [188]. Upon appearance of incapsulated lesions, at the age of 10 to 16 months, each nodule or liver tissue sample from KO animals was divided into three portions, one being snap frozen for DNA/RNA extraction, a second one fixed in 4% formaldehyde and further processed for histologic analysis, and a third one finely chopped and fixed in 1% formaldehyde-PBS for 15 minutes for chromatin immunoprecipitation. As a non-tumoral counterpart from the each individual, kidneys were collected and frozen.

Mouse histology

Livers portions assigned to histological assessment were fixed in 4% formaldehyde overnight. The next day the sample was washed in 70% ethanol and submitted for paraffin embedding. 5 um sections were stained with hematoxilin/eosin, and submitted for inspection to a mouse pathologist (Enrico Radaelli, VIB Center for the Biology of Disease, KU Leuven Center for Human Genetics, Belgium). The histological composition of grossly detectable hepatic nodules, along with the composition of the non-tumoral surrounding tissues, was semi-quantitatively determined based on the classification criteria reported in [189].

Mouse immunohistochemistry

Anti-CD3 immunostains were performed as previously described [160], with minor modifications. Briefly, 5 um sections were de-waxed and re-hydrated through an ethanol scale, heated in 1mM EDTA pH 8.0 in a water bath at 95°C for 25 minutes for

antigen de-masking and left to cool down for 20 min. After 5 minutes treatment with 3% H_2O_2, slides were incubated with rat monoclonal biotinilated CD3 antibody (Serotec) diluted 1:50 in a blocking solution (2% BSA, 2% goat serum, 0.02% Tween20, in TBS 1x) overnight at 4°C, washed three times with TBS 1x, and developed with DAB (DAKO) for 20 minutes. Slides were finally counterstained with hematoxylin, de-hydrated through alcoholic scale and mounted with Eukitt.

DNA, RNA and cDNA preparation

DNA from frozen hepatocytes was obtained using the DNeasy Blood and Tissue Kit (Qiagen). Frozen tissue samples were homogenized with a dounce homogenizer or with GentleMACS Dissociator (Miltenyi Biotec), depending on the tissue volume, prior to column extraction. DNA from formalin-fixed, paraffin-embedded (FFPE) samples was purified with the AllPrep DNA/RNA FFPE kit (Qiagen). RNA was extracted in Trizol (Invitrogen) using the RNeasy Mini Kit (Qiagen). 0.5 ug of total RNA was used for cDNA synthesis (using the ImProm-II Reverse Transcriptase, Promega), and 1 µl of the obtained cDNA was generally used as template for qPCR expression analyses. Quantification was performed on Nanodrop, and quality was assessed on Bioanalyzer (Agilent).

Chromatin Immunoprecipitation (ChIP)

ChIP was carried out as previously described [190]. Briefly, 2.5 millions of perfused fixed hepatocytes, or 350 mg of liver/tumoral fixed tissue have been used for each ChIP. Chopped tissue samples were further homogenized with GentleMACS Dissociator (Miltenyi Biotec) prior to lysis. Hepatocytes or homogenized tissues were processed with a two-step lysis protocol for cellular and nuclear membranes disruption, followed by chromatin shearing by sonication. Each lysate was then immunoprecipitated with 5 ug of the following antibodies: H3K27Ac (Abcam, ab4729), H3K4me1 (Abcam, ab8895), or H3K4me3 (Active Motif, #39159).

Antibodies were prebound overnight to 100 ul of G protein-coupled paramagnetic beads (Dynabeads) in PBS/BSA 0.5%. Beads were then added to lysates, and incubation was allowed to proceed overnight. Beads were washed six times in a

modified RIPA buffer (50 mM HEPES pH 7.6, 500 mM LiCl, 1 mM EDTA, 1% NP-40, and 0.7% Na-deoxycholate) and once in TE containing 50 mM NaCl. DNA was eluted in TE containing 2% SDS and crosslinks reversed by incubation overnight at 65**°**C. DNA was then purified by Qiaquick columns (Qiagen) and quantified with PicoGreen (Invitrogen).

ChIP validation by qPCR has been done using 1 ul of purified DNA for amplification on an Applied Biosystems 7500 Fast Real-time PCR system (SYBR Green, Applied Biosystems). Primers used for ChIP-qPCR are in reported in **Table 2**.

ChIP DNA was prepared for HiSeq2000 sequencing following standard protocols, with a 36bp single end setting.

Type of analysis	Type of primers	Forward	Reverse
ChIP-qPCR	H3K4me1 positive control 1	AGGCTCACAGAACCCAGAGA	TCACAAAGCATGCTCTCCAC
	H3K4me1 positive control 2	AGATGTCAGTTTGGGGTTGC	GTTGTCAGAAGGGCAGGTTC
	H3K4me3 positive control 1	AGAGTCAGGAGGGCAGGTTT	GCTCACCTCTTGGAGCATGT
	H3K4me3 positive control 2	AACTCTTGGCTTCTCGGGCA	ACCCTCCGTACTCTCCTGTG
	H3K4me1/3 negative control	GCCAAAGTGGAGTGGAAAGA	GCAGGTTCTGGAAACTGGAA
Expression	MKK7	AGGATCGACCTCAACTTGGA	GCTCTCTGAGGATGGTGAGC
	Nucleolin	TGGAGATCACAACAGCCAAA	CTTCTTCCGTTTTCCAGGTG

Table 2

List of primers used in this work. Type of analysis for which each primer couple has been used is indicated (top left column). H3K4me1 positive control primers have been designed on H3K4me1-rich regions according to published dataset (GEO accession n. GSM594586). H3K4me3 positive control primers have been designed around the TSS of genes expressed in liver (SCD1 and Cyp27A, respectively). H3K4me1/3 negative control primers have been designed on gene desert regions.

ChIPseq analyses

Short reads obtained either from Illumina HiSeq 2000 were quality-filtered according to the Illumina pipeline. Analysis of the data sets, beginning with alignment of the reads to the reference mm9 genome, was automated using the Fish the ChIPs pipeline [191]. All the reads with a unique match to the genome with two or fewer mismatches (-m 1 –v 2) were retained. Peak calling was performed using MACS v1.4 [192] with default parameters and bw = 300. Each IP was compared to input DNA derived from a WT liver.

To identify peaks increased in our inflamed or tumor samples datasets over the WT datasets, we used MACS [192].

Principal component analysis (PCA)

Principal component analysis is a statistical method that can be applied to a high-dimension dataset in order to reduce its dimensionality to a few original data dimensions: the principal components of the dataset [193]. In this case, PCA was used to reduce the dimensionality of tags number in all peaks (MACS score ≥50) in each samples considered to 2 main principal components, which accounts for a high percentage of the total variability. The number of tags has been normalized among samples prior to PCA, as samples had a different number of total starting tags (**Table S 1**).

Gene ontology analysis

For each timepoint sample group, GREAT 2.0.2 [194] was run against the whole mm9 genome (NCBI build 37 (UCSC mm9, Jul/2007)) as background, and results were summarized as a heatmap. We considered all the terms significantly enriched in at least one comparison (Hypergeometric FDR% ≤0.05) for the Biological Process ontology. FDR p-values were (-)log2-transformed in order to obtain positive scores. Results were then clustered using an hierarchical clustering algorithm (hclust R 2.15.1 function: method=average, Pearson correlation as a measure of the distance). Finally, scores were saturated (quantile = 0.95) and visualized (R heatmap.2 function).

Motifs enrichment analysis

In order to identify over-represented motifs corresponding to known TF binding sites, for H3K27Ac common induced peaks we used PSCAN [195] keeping H3K27Ac peaks that are unchanged in the different samples comparisons as background. Regions were scanned with 597 models (position weight matrices, PWM) collected from the literature [196-200].

To assess motif occurrence on the same datasets, we used Find individual motif occurrences (FIMO), a tool of the MEME Suit aimed at scanning a sequence database

for the occurrence of a desired list of motifs [201]. The list of matrices used as query for FIMO was derived from the top ranking PWM identified by PSCAN. We selected for occurrences with a minimum p-value of 10^{-4}, and searched for motifs at +/-500 bp around each peak center.

Target enrichment, whole exome, and whole genome sequencing

We performed target enrichment in order to enrich our genomic samples for desired sequences (in our case, the whole exome). This technique is based on the hybridization of the target sequences with commercial biotinylated RNA baits, selections of the hybrids with streptavidin and DNA fragments purification. For *Mdr2*-KO WES, SureSelect XT Mouse All Exon kit (Agilent) was used to target 21,543 mouse genes (**Table S 6**). As for human WES, target capture was done on six tumors and normal counterparts (**Table S 10**) using the SureSelect XT Human All Exon V4 kit (Agilent) targeting 20,965 human genes. Sample MB was excluded from whole exome sequencing because of the low tumor content (**Table S 9**).

Target enrichment was performed following the manufacturer's protocol with slight modifications. Briefly, around 3 ug of genomic DNA were sheared using an ultrasonic disruptor (Bioruptor, Diagenode) or using Adaptive Focused Acoustics technology (Covaris). After library preparation with Illumina DNA Sample Prep Kit, 200-250 bp fragments were selected and purified by gel extraction, or using the minelute PCR purification kit (Qiagen), or the Agencourt AMPure PCR Purification system (Beckman Coulter). Fragments were further amplified with 10 cycles of PCR and 500 ng were hybridized with each bait library. DNA capture was followed by paired-read cluster generation on the Cluster Station (Illumina). Libraries were sequenced using half lane of Illumina HiSeq2000 per sample, with 76 bp or 101 bp paired-end protocol, except for the tumoral sample of patient 7860, where one entire lane was used due to high levels of DNA degradation (**Table S 9**).

For whole genome sequencing, around 1 ug of mouse genomic DNA was sheared in 400-500 bp fragments using an ultrasonic disruptor (Bioruptor, Diagenode). Library was prepared with Illumina Paired-End DNA Sample Prep Kit. The libraries obtained

were sequenced using one lane of Illumina HiSeq2000 per sample, with the 101 bp paired-end protocol. Details on sequencing setting and throughput are available in **Table S 6**.

Alignment, variant calling and mutation validation

Paired-end sequencing reads from each tumor and normal were mapped to the mouse genome (NCBI37/mm9) or to the human genome (GRCh37/hg19) using Novoalign (http://novocraft.com). A maximum of three mismatches per read were allowed and duplicated reads were removed using rmdup of SAMtools [202]. All reads uniquely mapping within 75-100 bp of the targeted regions were considered on target and retained for further analysis.

Somatic point mutations are estimated to be rare events (roughly 1 per megabase), which imposes the requirement for a very low technical error rate to minimize false positive calls. On the other hand, tumor heterogeneity and sample contaminations with non-tumoral (stromal) cells might be frequent sources of false negatives [203]. In order to increase the sensitivity and specificity of mutation calling, tumor sequencing needs to reach a degree of coverage (i.e., the "covering" of the reference genome by the sequences generated from the analysed samples) of around 15-30x in whole genome experiments and 100x for whole exomes. Details on the coverage of our samples are found in **Table S 6** and **Table S 10**.

Single nucleotide variants (SNVs) and indels were identified using SAMtools [202] and Varscan 2 [204] and retained if covered by at least 10 reads and with frequency ≥20%. Somatic mutations and indels were identified as tumor-specific mutations with coverage ≥5x and frequency <10% in the normal counterpart. All retained SNVs and indels underwent manual inspection.

Genomic regions surrounding 26 random mutations were amplified by PCR using Taq DNA Polymerase (Qiagen) in the tumor and corresponding normal and Sanger sequenced in both directions on a 3730xl DNA Analyzer (Applied Biosystems) using dRhodamine chemistry. Out of 26 randomly selected variants, 25 were confirmed (> 96% accuracy).

CNV calling in mouse WES and WGS data

To detect CNVs on WES data, we exploited the difference in sequencing coverage between tumors and normal counterparts, based on the assumption that amplified genes have higher coverage in the tumors compared to their normal counterpart, whereas deleted genes have lower coverage (Sinha *et al*., submitted). After coverage normalization, the $\log_2$ratio of the fold change between tumor and normal was measured for each gene and used to assess the differences in copy number, setting as thresholds for amplification $\log_2$ratio = 0.5 (corresponding to >1.5 fold change), and for deletion $\log_2$ratio = -1.0 (corresponding to <0.5 fold change), respectively. In order to remove possible false positives due to particularly low coverage of some genes, we used PCA. Assuming a normal distribution of the variation in coverage between tumor and normal, we estimated a confidence interval based on the comparison between the CNVs detected in the whole genome and in the whole exome of the sample 60400/1. The optimal value that minimized false positive and maximized true positive was 90% confidence interval. Genes at the left tail of the distribution (above 90% PCA confidence interval) and with $\log_2$ratio ≥ 0.5 were considered as amplified, while genes at the right tail of the distribution (below 90% PCA confidence interval) and $\log_2$ratio ≤-1.0 were considered as deleted. CNVs were counted only for genes in the autosomal chromosomes. The genes in chromosome X were used as internal control for assessing amplifications while comparing female lesions with the normal male sample.

Copy number analysis on the WGS data was performed using CNVnator v. 0.2.5 [205] with sequence bins of 300bp. Regions with low confidence (p-value <0.05), composed of ≥65% repeats and/or gap, or covered for ≤30% of high quality mapping reads were removed. Furthermore, CNVs were retained only if their length spanning at least 3 consecutive bins was >1000bp. CNVs were identified as somatic if there was ≤ 5% overlap in length between tumor copy number regions and matched normal. To identify the CNV boundary, adjacent regions with the same copy number state were merged.

TaqMan copy number assay

Validation of putative copy number variations was performed by qPCR using TaqMan Copy Number Assay on a 7900HT Fast Real-Time PCR System (Applied Biosystems) with Sequence Detection Systems Software 2.2.2. TaqMan probes designed by the manufacturer were used for the experiment. Tert (Applied Biosystems, part number 4458373) was used as a reference. All samples were plated in quadruplicates with approximately 20 ng of DNA for each reaction. Copy-number calling was done with CopyCaller v2.0 (Applied Biosystems), using the matched normal tissue as a reference.

Statistical and network analyses

The probability that *Map2k7* amplification occurred at frequency higher than expected was computed by binomial test, using the fraction of all genes undergoing CNVs (2,430, 11% of the total targeted genes) as expectation.

Human protein-protein interaction network (13,531 proteins and 98,492 binary interactions) and the primary interactors of JNK (14 proteins) were retrieved from NCG3.0 [206]. 1,782 orthologs in the human network of the 3,095 mouse genes that were overall altered in the 14 analysed tumors were identified using MGI [207] and eggNOG 2.0 [208]. The distance of these proteins from the 14 JNK primary interactors was calculated using the function *shortest.paths* of the R IGRAPH package version 0.6 (http://igraph.sourceforge.net). Fourteen genes with the same degree (number of connections) of the JNK primary interactors were randomly extracted from the network for 10,000 times and their distance to the 1,782 proteins was measured to obtain a distribution of expected distances. Observed and expected distributions were compared using the Wilcoxon test.

Treatment with SP600125 JNK inhibitor

Twenty-three *Mdr2*-KO mice, 13 to 14 months old, were randomly divided into two groups and treated with SP600125 (anthra(1,9-cd)pyrazol-6(2H)-one) (Calbiochem), or vehicle. Vehicle for SP600125, diluted in DMSO, was 40% polyethylene glycol (PEG, Sigma) in PBS. Treatments (60 mg per dose) were administered intraperitoneally 3

times a week for a total of 3 weeks. Mice were sacrificed one week after the end of the treatment, and all detectable nodules were collected for DNA extraction and histological analysis. Grossly detectable hepatic nodules were counted and measured with a caliper.

Human sample description

Samples used in the study were obtained from frozen or FFPE material from 7 children diagnosed with PFIC2-related HCC (**Table S 9**). All specimens were obtained at native-liver hepatectomy during transplantation. Background liver in all patients exhibited parenchymal rather than portal-tract cholestasis, with BSEP expression detectable in none. Some patients displayed frank cirrhosis, others only variable degrees of fibrosis (**Table S 9**). Samples 7860 175, 1790, 2896, and MB came from single uncapsulated masses, while sample 23836 derived from one of several HCC within a single liver. Sample HB4R was an intrahepatic relapse that developed 6 years after liver transplantation, outside the graft. The patient was treated with chemotherapy before relapse surgical resection. In all samples, non-neoplastic liver tissues from the same patients were used as matching normal references.

Functional annotation of mutated genes

The list of genes affected by mutations was intersected with the genes know to be recurrently mutated in HCC (*TP53*, *CTNNB1*, *ARID1A*, *ARID2*, *AXIN1*, *PRS6KA3*, *VCAM1*, *CDK14*, *TERT*, *MLL4*, *CCNE1*) [209] and with the list of 537 genes known to have a causative role in human cancer [210].

Expression levels of mutated genes in normal liver were inferred from publicly available data [211, 212]. Starting from the raw CEL files of the two experiments (GSE2361 and GSE1133), data were normalized and analysed using the MAS5 algorithm [213]. The expression levels for each gene in the liver was calculated as the mean value between all gene probes with detection p-value <0.05. If all probes of a gene had p-value >0.05, the gene was considered not expressed. The normalized expression level was then measured as the gene expression level over the median expression of all genes in the liver. Genes with expression higher than the median

were considered as highly expressed, while all genes with expression lower than the median were defined as lowly expressed.

SNP array and copy number calling in the human samples

Genomic DNA extracted from FFPE samples and from frozen samples was processed according to Infinium® HD assay ultra manual. DNA from FFPE samples was restored before SNP array processing according to Infinium HD FFPE restore protocol. All seven human tumors and matched normal were assayed using Illumina HumanOmniExpress-12 v1.0 and image data were scanned using BeadArray reader. Intensity and genotype data were extracted for copy number variation analysis after normalizing raw fluorescent signals using Illumina Genome Studio.

Copy number variation analysis was performed using ASCAT (version 2.1) [214].

To identify amplified and deleted genes, the genomic coordinates of the CNV regions in each sample were intersected with those of 20,965 human genes contained in the SureSelect XT Human All Exon V4 kit (Agilent). A gene was considered as amplified or deleted if ≥ 80% of its length was contained in a CNV region.

Fluorescence *in situ* hybridization (FISH)

Validation of amplification of chromosome 19 in sample 23836 has been performed by two-color FISH using a Vysis LSI 19q13 SpectrumOrange/ 19p13 SpectrumGreen probe (Abbott), according to manufacturer's instructions. 2 um FFPE slides from tumor and normal liver of patient 23286 were first deparaffinised in xylene, washed in 100% ethanol, incubated in 1x SSC (0.3 M sodium chloride, 0.03 M sodium citrate) pH 6.0 at 80°C for 20 minutes for de-masking, and digested with pepsin (0.5 mg/ml in 0.2 N HCl, pH 1.0; Protease and Protease Buffer II, Abbott) for 17 minutes at 37°C. Samples were then washed in 2x SSC, dehydrated in 70, 95 and 100% ethanol, and air dried. 10 ul of probe were directly applied on each slide, topped with a coverglass, and sealed with rubber cement. Slides were placed in a HYBrite (Abbott), and the probe was left to denature 1 minute at 85°C, followed by an overnight hybridization at 37°C. Coverglasses were finally removed and slides were washed twice in 2x SSC with 0.1% NP-40 at RT, once in 0.4x SSC with 0.3% NP-40 at 73°C, and once again in 2x SSC

with 0.1% NP-40 at RT. After counterstaining with DAPI (Sigma), FISH signals were scored with an Olympus BX61 upright microscope, using a 100x objective.

RESULTS

1. Epigenomic analyses

1.1 Histological characterization of *Mdr2*-WT and –KO samples

In order to identify possible changes in epigenomic patterns along the *Mdr2*-KO tumorigenic process, we generated a panel of chromatin immunoprecipitations (ChIP) followed by NGS (ChIPseq) on liver samples prepared from different time points. In particular, we considered at least 3 samples of adenoma, low grade HCC (≤60% carcinoma content) and high grade HCC (≥80% carcinoma content), all coming from different individuals 11 to 16 months old. These samples have been compared to age-matched WT livers, and to samples of 8 months old inflamed, non-tumoral livers (**Table 3**).

Pathological inspection of the selected samples identified an entire spectrum of histological situations. From the baseline, represented by normal hepatic parenchyma without any microscopic lesion, the disease starts from the occurrence of adaptive, degenerative or reactive features of hepatocytes in response to an inflammatory condition. These changes include hepatocellular atrophy, regenerative hepatocellular hyperplasia, hyperplasia of oval cells or bile ducts, hepatocellular degeneration or necrosis, inflammatory or fibroproliferative changes. Progression leads to the development of foci of preneoplastic cellular alteration, non-incapsulated but discrete lesions demarcated from the adjacent tissue by virtue of different (generally increased) cell size and peculiar staining (which classifies them as basophilic, eosinophilic, amphophilic, clear cell or vacuolated). Adenomas are distinguished as incapsulated and often bigger (and macroscopically visible) lesions, with slight signs of cytologic atypia (coarsely clumped chromatin, large nucleoli, increased nucleus to cytoplasm ratio, cytoplasmic basophilia). The last stage of the disease is characterized by the appearance of carcinoma, which could display peculiar features such as vascularization, aberrant mitotic figures, moderate to marked cytologic atypia, internal necrosis or hemorrage and typical cellular architecture (with trabecular, acinar or solid patterns) (**Table 3**).

Type	Sample	Gender	Preparation	% normal	% degenerative changes	% preneoplastic foci	% adenoma	% carcinoma
WT	1	F	Perfusion	N.A.	N.A.	N.A.	N.A.	N.A.
	2	M	Perfusion	N.A.	N.A.	N.A.	N.A.	N.A.
	3	M	Perfusion	N.A.	N.A.	N.A.	N.A.	N.A.
	4	M	Tissue	100	0	0	0	0
	Pool (samples 1, 2, 3, 4)		N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Inflamed	1	M	Tissue	0	90	10	0	0
	2	F	Perfusion	N.A.	N.A.	N.A.	N.A.	N.A.
	3	F	Perfusion	N.A.	N.A.	N.A.	N.A.	N.A.
	4	F	Perfusion	N.A.	N.A.	N.A.	N.A.	N.A.
Adenoma	1	F	Tissue	0	0	0	100	0
	2	F	Tissue	0	0	0	100	0
	3	F	Tissue	0	20	10	70	0
	4	F	Tissue	0	20	10	70	0
HCC low grade	1	M	Tissue	0	30	10	50	10
	2	F	Tissue	0	30	10	50	20
	3	M	Tissue	0	30	0	40	20
	4	F	Tissue	0	20	10	30	40
	5	F	Tissue	0	10	0	30	60
HCC high grade	1	F	Tissue	0	10	0	0	90
	2	F	Tissue	0	10	0	0	90
	3	M	Tissue	0	0	0	20	80
	Pool (samples 1, 2, 3)		Tissue	N.A.	N.A.	N.A.	N.A.	N.A.

Table 3

Histologic description of samples processed for ChIPseq analyses. Hepatocyte preparation via collagenase perfusion is not compatible with material being processed for histologic examination, therefore for perfused samples histologic composition was not assessed. One sample of tissue processed without perfusion has been included in each set, in order to verify whether possible differences could be ascribed to the preparation protocol rather than to the histological content. In the case the degree of variability (dependent on the fact we are considering different tissues from different individuals) would have been too high to allow a proper statistical separation among different categories, we included a pool of all WT and all high grade HCC samples. N.A. = not applicable.

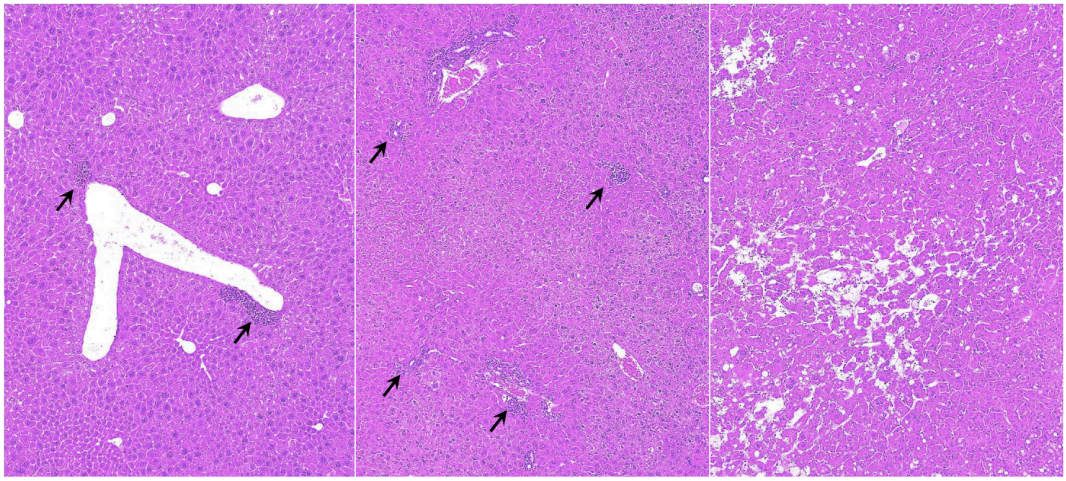
Hepatocellular carcinomas develop within adenomas, implying that each nodule is a mixture of adenoma and carcinoma content, whose proportion has been carefully evaluated. Furthermore, in order to estimate local invasiveness (one of the most reliable indicator of tumor malignancy), it is necessary to pathologically inspect also the tissue surrounding the lesions (i.e. normal, inflamed or pre-neoplastic tissue). For this reason, each nodular specimen assigned to histologic analysis includes a non-tumoral portion (**Table 3**, **Figure 4**). As for the genetic and epigenetic analyses, however, in order to avoid confounding results due to contamination by surrounding inflamed, non-tumoral tissue, each incapsulated lesion (adenoma or HCC) has been carefully excised, cleaned from surrounding material and individually processed.

A

Class	Sample n.	Normal	Degenerative changes	Preneoplastic foci	Adenoma	Carcinoma
WT	4	100% -	0%	0%	0%	0%
Inflamed	1	0%	90% +	10% -	0%	0%

Class	Sample n.	Surrounding, external tissue			Incapsulated lesions	
		Normal	Degenerative changes	Preneoplastic foci	Adenoma	Carcinoma
Adenoma	1	0%	0%	0%	100% -	0%
	2	0%	0%	0%	100% -	0%
	3	0%	20% +	10% -	70% -	0%
	4	0%	20% ++	10% -	70% -	0%
Low grade HCC	1	0%	30% +	10% -	50% -	10% -
	2	0%	30% ++	0%	50% -	20% -
	3	0%	30% -	10% -	40% -	20% -
	4	0%	20% -	10% -	30% -	40% -
	5	0%	10% -	0%	30% -	60% -
High grade HCC	1	0%	10% +	0%	0%	90% -
	2	0%	10% +	0%	0%	90% -
	3	0%	0%	0%	20% -	80% -

B



WT sample 4
normal

Inflamed sample 1
multifocal mild chronic cholangiohepatitis

Low grade HCC sample 5
hepatocellular carcinoma

Figure 4

Histologic assessment of inflammatory infiltrates presence in parenchymal liver tissues along *Mdr2*-KO pathogenesis.

(A) Semi-quantitative score of inflammatory changes observed in the tissues used for ChIPseq. Changes were characterized by focal infiltrates of inflammatory cells including one or more of the following cell types: lymphocytes, plasma cells, macrophages (often pigmented) and

neutrophils. Hematoxilin/eosin sections were scanned in at 10x magnification and for each histologic component (expressed as a percentage) identified in the sample the degree of inflammatory changes was indicated as – (average number of inflammatory foci ≤ 2 per miscoscopic 10x field), + (average number of inflammatory foci > 2 and ≤ 5 per miscoscopic 10x field), ++ (average number of inflammatory foci > 5 per miscoscopic 10x field).
(B) Representative microscopic fields of some analyzed specimens. All pictures have been captured at 10x magnification. Arrows denote inflammatory cell foci.

In the case of WT and inflamed organs, where the whole liver had to be taken but contamination could result from non-parenchymal cell populations (e.g. Kupffer cells, endothelial cells), collagenase perfusion has been applied to obtain a single-cell suspension that has been purified by centrifugation and thereby enriched for hepatocytes (**Fig. S4**).
On the other hand, inside nodular formations (adenoma and carcinoma), pathological examination estimated a low degree of inflammatory cells infiltration, which was comparable to the one found in the livers of healthy individuals (**Figure 4**).

1.2 Statistical analysis of the variability among *Mdr2* ChIPseq datasets

Our primary interest was to identify transcription factors (TFs) becoming activated during the tumorigenic process, and that could thus provide hints on the principal mechanisms accompanying the generation of cancer. For this reason we chose to follow a "reverse epigenetics" approach, by focusing on the profiling of H3K27Ac modification as a mark of active regions, that could suggest TF potentially relevant for disease progression [215]. To this purpose, we performed a set of ChIPseqs against H3K27Ac on inflamed, adenoma, low grade and high-grade HCC samples, to be compared with ChIPseqs on WT samples. Furthermore, we were interested in searching for changes at the level of regulatory regions such as active promoters or distal elements (enhancers). For this reason, we generated ChIPseq datasets to profile H3K4me3 (a mark for active transcription start sites, TSS) and H3K4me1 (enriched on enhancers) on WT, inflamed and low-grade HCC samples [216].

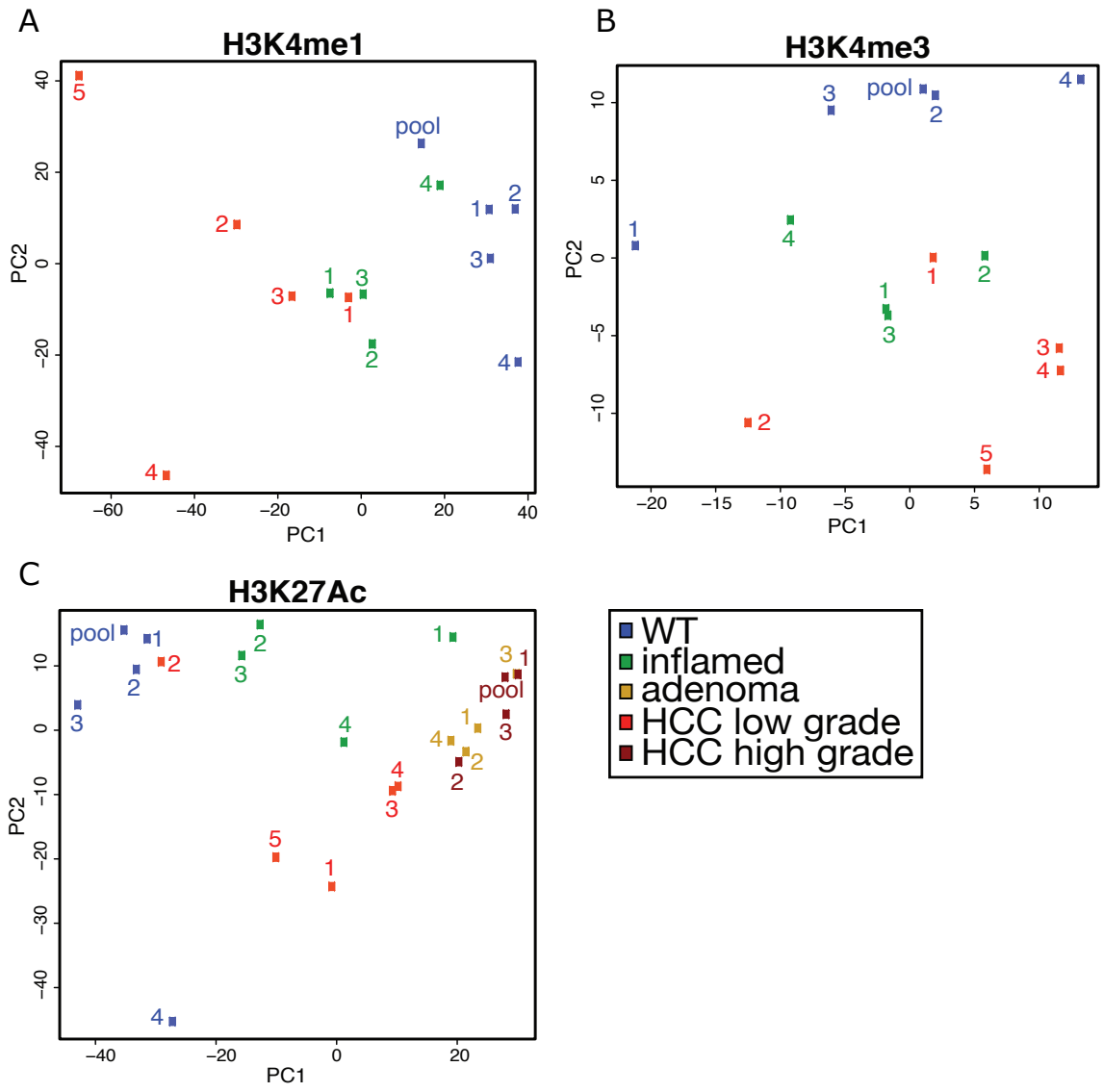


Figure 5

Principal component analysis of *Mdr2*-KO ChIPseq data sets.

As a first step, we wanted to check whether the degree of variability among samples was too high to allow discrimination of different disease stages. In order to reduce the complexity of these data sets, we applied a principal component analysis (PCA), a statistical procedure aimed at extrapolating a number of dimensions as small as possible that would describe the original data set while maintaining much of its variance. In particular, PCA converts a number of possibly correlated variables into a number of uncorrelated variables (or principal components, PCs) by orthogonal transformation. In this system, the first PC includes as much variability as possible (and therefore, it is usually the most informative parameter), and each subsequent PC

describes as much of the remaining variability as possible (provided it is orthogonal to the preceding PCs) [193]. Usually, the first two PC provide a simplified but sufficiently informative view of the data.

In this case, PCA was used to reduce the dimensionality of tags number to 2 main components. By applying a PCA to the ChIPseq data for H3K4me1 and H3K4me3, we observed that in both cases samples belonging to different disease steps were generally clustering separately (**Figure 5B**). In the case of H3K27Ac, for which data for two additional categories ("adenoma" and "high grade HCC") was available, separation was still generally neat (**Figure 5C**).

Based on PCA results, for further analyses we excluded as "outliers" samples failing to cluster close to the bulk of each category.

1.3 Assessment of common H3K27Ac changes at each timepoint along *Mdr2* tumorigenic process

In order to find epigenomic changes characteristic of each disease stage, for each modification that was considered we calculated the number of induced or repressed peaks over the WT that were in common to all samples of a given group (**Table 4**). The highest number of induced or repressed peaks (with respect to WT) displayed by all samples at a given time point was found for H3K27Ac, and particularly for induced peaks, indicating that this mark is subject to the highest variability specifically associated to the *Mdr2*-KO tumorigenic process. On the contrary, the number of common induced or repressed H3K4me1 and me3 peaks is very low both for inflamed and low grade HCC samples.

These data suggests that H3K27Ac-induced regions are possibly most informative on the *Mdr2*-KO pathogenesis, and for this reason we chose to focus further analyses on this modification.

	H3K27Ac		H3K4me1		H3K4me3	
	Induced	Repressed	Induced	Repressed	Induced	Repressed
Inflamed	1608	78	546	66	2	23
Adenoma	2010	257	N.A.	N.A.	N.A.	N.A.
Low grade HCC	1408	114	409	172	102	77
High grade HCC	2306	501	N.A.	N.A.	N.A.	N.A.

Table 4

Number of induced or repressed H3K27Ac, H3K4me1 or H3K4me3 ChIPseq peaks that are in common to each disease timepoint. N.A. = not applicable.

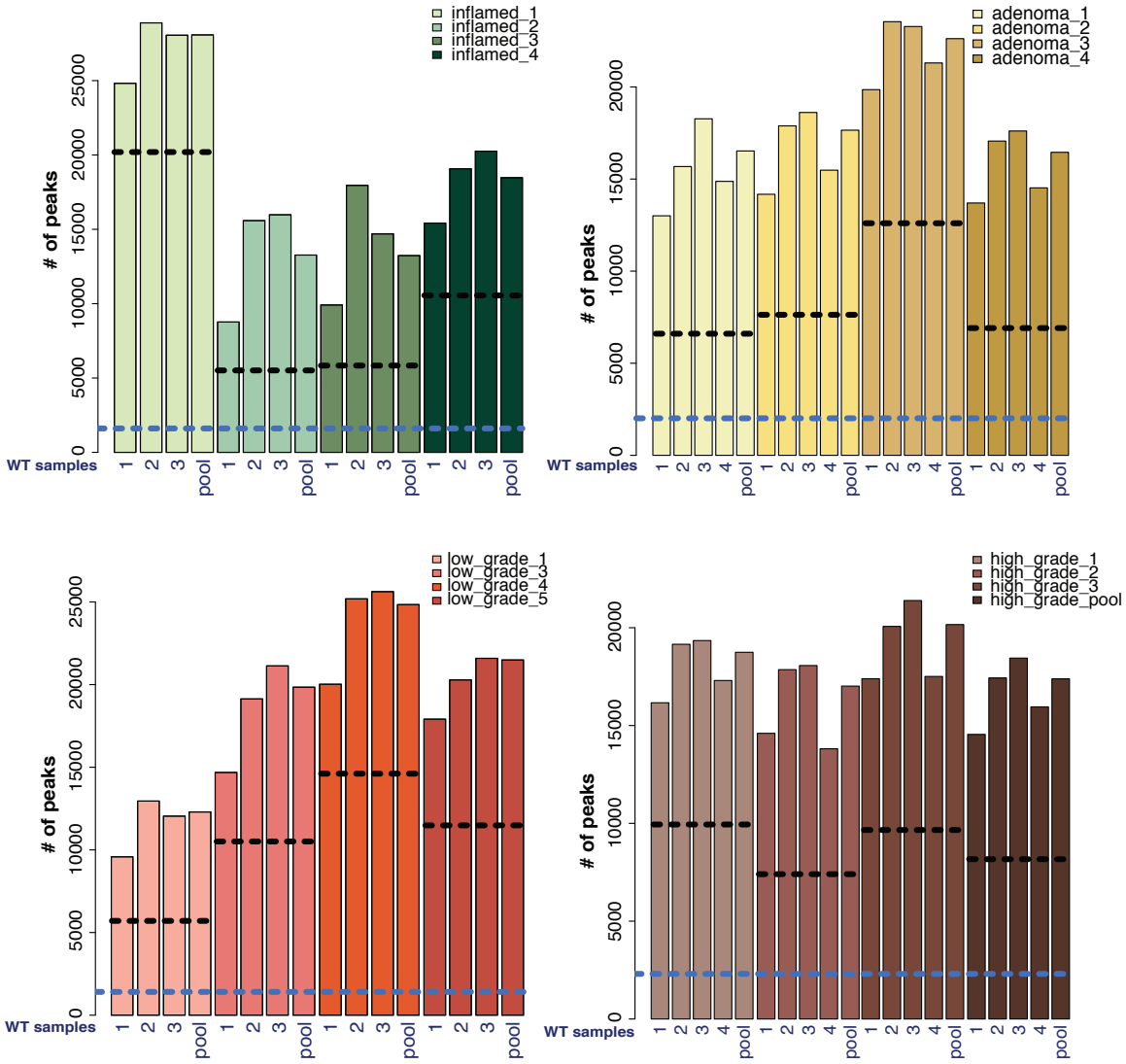


Figure 6

Number of peaks for *Mdr2*-KO H3K27Ac ChIPseq samples. Numbers of WT samples (in blue) are reported at the bottom of each bar. Dashed black lines indicate the number of common peaks of each sample when compared to each of the WT samples, while dashed blue lines indicate common peaks among all samples compared to all WTs.

Taking a closer look, by normalizing the number of peaks of each sample with the number of peaks of each WT sample, generally half of the peaks are common to all 4 comparisons with WTs suggesting that, as expected, WT samples are characterized by a low degree of variability (**Figure 6**). However, the number of peaks common to all the comparisons is much lower, underlying the high variability in the evolution of liver cancer. Nevertheless, peaks found to be common to all the comparisons within each category can be considered as stage-specific peaks, and were therefore taken in consideration for downstream analyses. In particular, as previously discussed (**Table 4**), we focused on H3K27Ac induced peaks that were common to all samples in each timepoint.

In order to evaluate whether enhancers that became hyperacetylated with respect to the WT (i.e., enhancers that became active) also displayed changes in H3K4me1, we analyzed the H3K4me1 status on the common regions that acquired H3K27Ac in inflamed or low grade HCC samples (**Figure 7**). Both in the case of inflamed and low grade HCC timepoints, we observed that regions gaining H3K27Ac were undergoing changes in methylation levels, and in particular they were generally acquiring H3K4me1 (**Figure 7**).

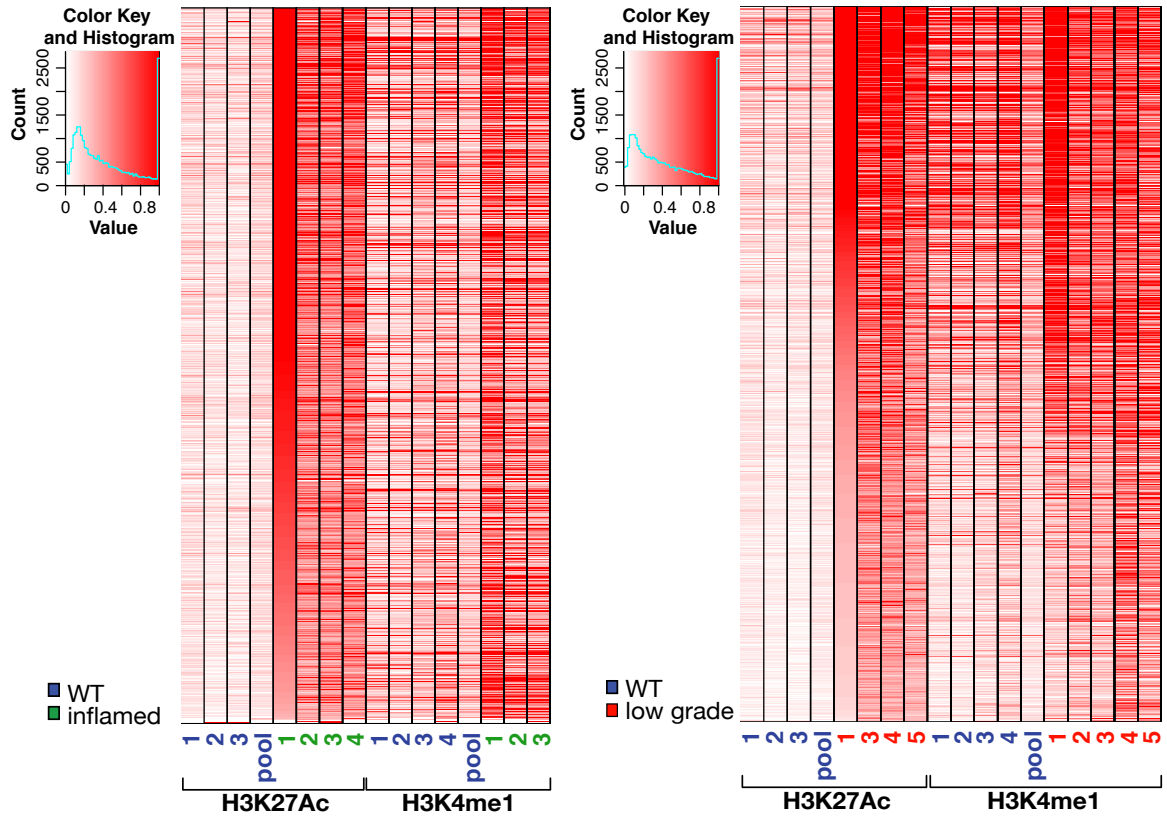


Figure 7

H3K4me1 status at H3K27Ac common induced ChIPseq peaks in *Mdr2*-KO inflamed and low grade HCC samples. Regions associated to H3K27Ac common induced peaks were observed in H3K27Ac and, in parallel, in H3K4me1 ChIPseq samples, as indicated on bottom. Rows were ordered by decreasing number of tags according to the first inflamed (left heatmap) or HCC (right heatmap) sample.

1.4 Gene ontology analysis on H3K27Ac common induced peaks-associated genes

We next wanted to score H3K27Ac common induced peaks for associated functional categories of the nearby genes. To this end, we performed a gene ontology (GO) analysis using the "genomic regions enrichment of annotations tool" (GREAT) [194], using WT samples as reference. GREAT failed to cluster samples according to the tumorigenesis timepoint classes (**Figure 8**). However, a strong inter-class functional enrichment was observed for biological processes associated to inflammation and immune response (**Figure 8**). This suggests an activation of machineries involved in immunity and inflammatory components, which would start from the inflammation

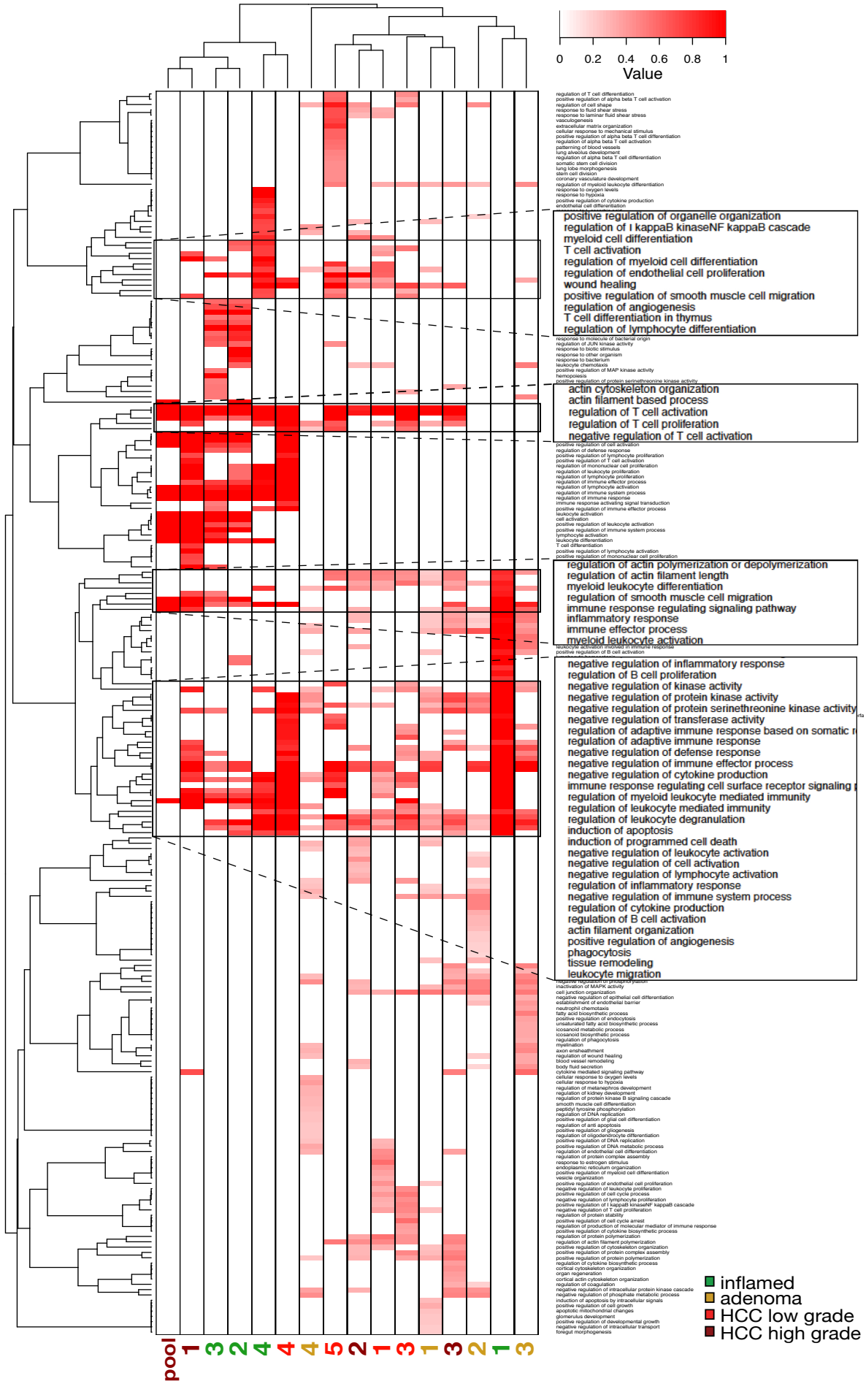

0
0.2
0.4
0.6
0.8
1
Value
positive regulation of organelle organization
regulation of I kappaB kinaseNF kappaB cascade
myeloid cell differentiation
T cell activation
regulation of myeloid cell differentiation
regulation of endothelial cell proliferation
wound healing
positive regulation of smooth muscle cell migration
regulation of angiogenesis
T cell differentiation in thymus
regulation of lymphocyte differentiation
actin cytoskeleton organization
actin filament based process
regulation of T cell activation
regulation of T cell proliferation
negative regulation of T cell activation
regulation of actin polymerization or depolymerization
regulation of actin filament length
myeloid leukocyte differentiation
regulation of smooth muscle cell migration
immune response regulating signaling pathway
inflammatory response
immune effector process
myeloid leukocyte activation
negative regulation of inflammatory response
regulation of B cell proliferation
negative regulation of kinase activity
negative regulation of protein kinase activity
negative regulation of protein serinethreonine kinase activity
negative regulation of transferase activity
regulation of adaptive immune response based on somatic r
regulation of adaptive immune response
negative regulation of defense response
negative regulation of immune effector process
negative regulation of cytokine production
immune response regulating cell surface receptor signaling
regulation of myeloid leukocyte mediated immunity
regulation of leukocyte mediated immunity
regulation of leukocyte degranulation
induction of apoptosis
induction of programmed cell death
negative regulation of leukocyte activation
negative regulation of cell activation
negative regulation of lymphocyte activation
regulation of inflammatory response
negative regulation of immune system process
regulation of cytokine production
regulation of B cell activation
actin filament organization
positive regulation of angiogenesis
phagocytosis
tissue remodeling
leukocyte migration
pool
1 3 2 4 4 4 5 2 1 3 1 3 2 1 3
inflamed
adenoma
HCC low grade
HCC high grade

Figure 8

GREAT analysis of the biological processes linked to genes associated with H3K27Ac gain in *Mdr2*-KO ChIPseq samples. Close-ups inserts highlight affected categories that are common to the majority of samples.

phase and would be retained until end-stage tumor, despite the absence of non-parenchymal, inflammatory cells in tumoral samples (**Figure 4**).

1.5 Motifs enrichment analysis on H3K27Ac common induced peaks

In order to identify over-represented motifs representing known TF binding sites in H3K27Ac common induced peaks, we used PSCAN [195]. H3K27Ac peaks that did not vary along the different samples were used as background for comparison.

At all time points, the most enriched matrices corresponded to binding sites for the activator protein 1 (AP-1) TF family members (such as Jun, JunB, JunD, JDP2, Fos, FosB; **Table 5**). AP-1 TFs are heterodimers composed of members of the Fos (FBJ murine osteosarcoma viral oncogene homolog), Jun, JDP (Jun-dimerizing partner) or ATF (activating transcription factor) subfamilies [217]. They are activated in response to a variety of stimuli, including cell stress and inflammation, exposure to cytokines or growth factors and infection, and control cell proliferation and apoptosis [218]. Therefore, the finding that common H3K27Ac induced peaks are enriched for AP-1 matrices is in agreement with the general activation of inflammatory pathways resulting from the GO analysis (**Figure 8**).

The second class of transcription factors whose matrices were enriched in H3K27Ac common induced peaks was the erythroblast transformation-specific (ETS) TF family. Predominant ETS matrices were the ones associated to the SPI subfamily (from spleen focus forming virus (SFFV) proviral integration oncogene), namely Spi1 (also called PU.1), SpiB and SpiC (**Table 5**). ETS members are involved in the regulation of a number of key cellular processes, such as proliferation, differentiation, cell cycle, apoptosis and transformation [219]. Furthermore, they are critical for hematopoiesis and immune response [220].

Inflamed		
TF	**Matrix ID**	**P-value**
JDP2	TA0409_JDP2_full_dimer	1.37E-114
JDP2	TA0411_Jdp2_DBD_dimer	4.08E-112
JDP2	TA0407_JDP2_DBD_dimer	5.34E-108
NFE2	TA0422_NFE2_DBD_dimer	5.14E-103
JunD	HC_JUND_f1	1.94E-97
BATF	HC_BATF_si	1.96E-92
Jun	HC_JUN_f1	3.17E-91
FosL2	HC_FOSL2_f1	7.15E-91
JunB	HC_JUNB_f1	5.34E-89
SMRC1	HC_SMRC1_f1	1.66E-76
FosB	HC_FOSB_f1	2.00E-74
PU.1	HC_SPI1_si	7.54E-69
Fos	MA0099.1 Fos	1.43E-64
AP1	MA0099.2 AP1	5.56E-63
PU.1	TA0154_SPIB_DBD_monomer	2.01E-54
PU.1	BU0058 Sfpi1_primary	2.72E-53
IRF8	GN_IRF8_TOP5000	2.95E-48
SPIC	TA0156_Spic_DBD_monomer	5.21E-48
SPIB	HC_SPIB_f1	1.79E-45
PU.1	TA0153_SPI1_full_monomer	7.99E-41
FosL1	HC_FOSL1_f2	3.78E-40
PU.1	MA0080.2 SPI1	7.51E-37
SPIC	TA0155_SPIC_full_monomer	1.91E-36
ELF5	HC_ELF5_f1	2.26E-35
PU.1	ETS0027 h-SPI1	2.47E-35

Adenoma		
TF	**Matrix ID**	**P-value**
FosL2	HC_FOSL2_f1	3.64E-89
JunD	HC_JUND_f1	1.36E-83
Jun	HC_JUN_f1	1.46E-83
SMRC1	HC_SMRC1_f1	2.22E-76
JunB	HC_JUNB_f1	2.82E-73
BATF	HC_BATF_si	1.04E-66
JDP2	TA0409_JDP2_full_dimer	2.98E-64
JDP2	TA0407_JDP2_DBD_dimer	2.25E-63
JDP2	TA0411_Jdp2_DBD_dimer	4.64E-62
NFE2	TA0422_NFE2_DBD_dimer	2.78E-56
Fos	MA0099.1 Fos	1.54E-53
FosB	HC_FOSB_f1	1.40E-50
AP1	MA0099.2 AP1	5.48E-46
FosL1	HC_FOSL1_f2	3.72E-31
PU.1	ETS0027 h-SPI1	7.46E-20
SPIB	ETS0025 h-SPIB	1.95E-19
NFE2	HC_NFE2_f2	5.30E-19
SPIC	ETS0026 h-SPIC	3.27E-18
BACH1	HC_BACH1_si	1.09E-17
PU.1	HC_SPI1_si	5.47E-17
MafK	HC_MAFK_si	1.16E-16
ETS2	HC_ETS2_f1	1.14E-14
ETV4	HC_ETV4_f1	2.15E-14
NRF2	NRF2_NAR	1.85E-13
Zfp281	BU0097 Zfp281_primary	1.43E-12

Low grade HCC		
TF	**Matrix ID**	**P-value**
JDP2	TA0409_JDP2_full_dimer	1.41E-43
JDP2	TA0407_JDP2_DBD_dimer	2.56E-43
JDP2	TA0411_Jdp2_DBD_dimer	2.11E-42
BATF	HC_BATF_si	1.99E-39
JunB	HC_JUNB_f1	5.31E-38
JunD	HC_JUND_f1	5.94E-38
NFE2	TA0422_NFE2_DBD_dimer	2.48E-37
FosB	HC_FOSB_f1	8.76E-36
Jun	HC_JUN_f1	1.90E-35
FosL2	HC_FOSL2_f1	7.82E-31
SMRC1	HC_SMRC1_f1	1.57E-30
Fos	MA0099.1 Fos	3.59E-30
AP1	MA0099.2 AP1	2.73E-27
TEF1	TA0317_TEAD1_full_monomer	8.60E-27
TEF5	TA0320_TEAD3_DBD_monomer	1.57E-22
PU.1	HC_SPI1_si	1.09E-21
GATA2	HC_GATA2_si	5.56E-21
TEF3	TA0321_TEAD4_DBD_monomer	1.20E-20
ELF5	MA0136.1 ELF5	3.15E-19
GATA3	TA0157_GATA3_DBD_monomer	1.13E-18
GATA3	TA0158_GATA3_full_monomer	9.10E-18
GATA4	TA0159_GATA4_DBD_monomer	1.50E-17
GATA3	HC_GATA3_si	2.12E-17
ELF5	HC_ELF5_f1	5.23E-17
GATA1	MA0035.2 Gata1	1.00E-16

High grade HCC		
TF	**Matrix ID**	**P-value**
FosL2	HC_FOSL2_f1	8.32E-103
JunD	HC_JUND_f1	1.85E-99
Jun	HC_JUN_f1	9.33E-97
SMRC1	HC_SMRC1_f1	3.18E-91
JDP2	TA0409_JDP2_full_dimer	6.50E-80
JDP2	TA0407_JDP2_DBD_dimer	1.67E-77
JunB	HC_JUNB_f1	1.26E-75
Fos	MA0099.1 Fos	1.88E-74
JDP2	TA0411_Jdp2_DBD_dimer	2.51E-73
BATF	HC_BATF_si	2.93E-72
NFE2	TA0422_NFE2_DBD_dimer	2.41E-70
FosB	HC_FOSB_f1	4.58E-67
PU.1	HC_SPI1_si	1.42E-56
AP1	MA0099.2 AP1	2.28E-51
FosL1	HC_FOSL1_f2	2.76E-40
PU.1	BU0058 Sfpi1_primary	6.83E-38
SPIB	ETS0025 h-SPIB	1.88E-36
PU.1	ETS0027 h-SPI1	1.53E-35
SPIC	ETS0026 h-SPIC	7.73E-35
SPIB	TA0154_SPIB_DBD_monomer	2.80E-29
IRF8	GN_IRF8_TOP5000	4.68E-25
SPIB	HC_SPIB_f1	1.51E-22
ETV4	HC_ETV4_f1	1.69E-22
SPIC	TA0156_Spic_DBD_monomer	2.02E-21
NFE2	HC_NFE2_f2	3.77E-21

Table 5

Top 25 enriched motifs on H3K27Ac common induced ChIPseq peaks as ranked by PSCAN. TFs known to recognize each motif are indicated.

We next wanted to assess how many AP1 and ETS matrices were present in the H3K27Ac common induced peaks. By submitting the matrices that were most often represented at the top of the PSCAN ranks to Find individual motif occurrences (FIMO), we found that AP1 matrices were present on ~30-35% of H3K27Ac common induced peaks, while ETS members had varying yet higher degrees of occurrences (PU.1 matrices being the most present) (**Table 6**). Also in this case there was no

striking difference among different timepoint classes.

		N. inflamed	%	N. adenoma	%	N. low grade	%	N. high grade	%
Total H3K27Ac common induced peaks		1608	100	2010	100	1408	100	2306	100
AP1 family	HC_JUN_f1-results-uniq.tab	616	38.3	775	38.6	459	32.6	897	38.9
	HC_JUNB_f1-results-uniq.tab	653	40.6	800	39.8	474	33.7	914	39.6
	HC_JUND_f1-results-uniq.tab	645	40.1	792	39.4	472	33.5	930	40.3
	TA0407_JDP2_DBD_dimer-results-uniq.tab	428	26.6	482	24.0	278	19.7	535	23.2
	TA0409_JDP2_full_dimer-results-uniq.tab	488	30.3	551	27.4	312	22.2	621	26.9
	TA0411_Jdp2_DBD_dimer-results-uniq.tab	456	28.4	515	25.6	290	20.6	578	25.1
	MA0099.1.Fos-results-uniq.tab	464	28.9	550	27.4	325	23.1	647	28.1
	HC_FOSB_f1-results-uniq.tab	586	36.4	712	35.4	460	32.7	844	36.6
	HC_FOSL1_f2-results-uniq.tab	638	39.7	805	40.0	511	36.3	943	40.9
	HC_FOSL2_f1-results-uniq.tab	620	38.6	765	38.1	444	31.5	899	39.0
	HC_NFE2_f2-results-uniq.tab	586	36.4	753	37.5	454	32.2	867	37.6
	TA0422_NFE2_DBD_dimer-results-uniq.tab	473	29.4	531	26.4	304	21.6	619	26.8
	HC_BATF_si-results-uniq.tab	630	39.2	790	39.3	467	33.2	887	38.5
	MA0099.2.AP1-results-uniq.tab	565	35.1	706	35.1	420	29.8	819	35.5
ETS family	HC_SPI1_si-results-uniq.tab	1112	69.2	1267	63.0	905	64.3	1562	67.7
	BU0058.Sfpi1_primary-results-uniq.tab	690	42.9	676	33.6	486	34.5	904	39.2
	MA0080.2.SPI1-results-uniq.tab	578	35.9	591	29.4	432	30.7	746	32.4
	ETS0027.h-SPI1-results-uniq.tab	700	43.5	747	37.2	497	35.3	929	40.3
	TA0153_SPI1_full_monomer-results-uniq.tab	713	44.3	662	32.9	513	36.4	888	38.5
	HC_SPIB_f1-results-uniq.tab	766	47.6	760	37.8	565	40.1	974	42.2
	ETS0025.h-SPIB-results-uniq.tab	667	41.5	751	37.4	499	35.4	929	40.3
	TA0154_SPIB_DBD_monomer-results-uniq.tab	747	46.5	705	35.1	512	36.4	944	40.9
	ETS0026.h-SPIC-results-uniq.tab	734	45.6	822	40.9	559	39.7	1012	43.9
	TA0155_SPIC_full_monomer-results-uniq.tab	636	39.6	635	31.6	474	33.7	805	34.9
	TA0156_Spic_DBD_monomer-results-uniq.tab	700	43.5	692	34.4	500	35.5	894	38.8
	HC_ELF2_f1-results-uniq.tab	835	51.9	1000	49.8	637	45.2	1172	50.8
	BU0012.Elf3_primary-results-uniq.tab	562	35.0	574	28.6	414	29.4	723	31.4
	HC_ELF5_f1-results-uniq.tab	531	33.0	567	28.2	419	29.8	678	29.4
	MA0136.1.ELF5-results-uniq.tab	582	36.2	630	31.3	461	32.7	760	33.0
	HC_ETS2_f1-results-uniq.tab	723	45.0	931	46.3	600	42.6	1058	45.9
	HC_ETV4_f1-results-uniq.tab	881	54.8	1027	51.1	686	48.7	1249	54.2
	HC_FLI1_f1-results-uniq.tab	700	43.5	917	45.6	555	39.4	1069	46.4

Table 6

Motifs occurrence analysis on H3K27Ac common induced ChIPseq peaks by FIMO. The number of occurrences, along with the relative percentage calculated on the total of the H3K27Ac common induced peaks for each stage, is shown. Motif occurrences with p-values smaller than 10^{-4} have been considered.

2. Genetic analyses[1]

2.1 Single nucleotide variations analyses on *Mdr2*-KO HCCs

The second part of this project was intended to screen for genetic modifications occurring during *Mdr2*-KO tumorigenesis. We aimed at identifying genetic alterations in tumors, in order to subsequently check which alteration is already appearing at earlier stages.

[1] This part is product of collaboration with F. Iannelli, S. Sinha and F. Ciccarelli, European Institute of Oncology (IEO), Milan.

To map the somatic mutations acquired in *Mdr2*-KO lesions, target enrichment coupled to NGS was used. Nine nodules with varying percentages of HCC content, extracted from 10 to 16 months old individuals, were compared to a non-tumoral tissue (kidney) from one of these mice (**Table 7**). Mutational analysis was performed using a mouse all-exome probes array, targeting 21,543 genes (for a total 50,4 Mbp). We obtained 22,4 Gb of raw reads per sample, with coverage ranging from 100 to 150x. After variant calling (see Materials and Metods), we identified a total of 118 SNVs, none of which was shared between any two tumors, and no indels (**Tables 7**), with less than 0.4% of false positives. Sixty SNVs (51%) on a total of 60 genes were non-silent.

Interestingly, *Mdr2*-KO lesions bore a lower number of somatic mutations per Mbp with respect to other solid tumors, including human HCC (**Figure 9A**). None of the 60 non-silent mutations affected any ortholog of the 11 genes most frequently altered in human HCC , or any other gene known to be causally implicated in cancer (data not shown). Furthermore, the length of genes associated to these mutations was significantly longer than for other mouse genes (**Figure 9B**), implying a higher statistical chance of random mutations in these genes. Additionally, 74% of mutated genes were either poorly or not expressed in mouse liver (**Figure 9C**) indicating that they may be not functional in this tissue. Taken together, these observations suggest that these non-silent mutations do not have a driver role, and that therefore mutational instability is not the driving force in the development of liver cancer in *Mdr2*-KO mice.

Lesion ID	Gender	% adenoma	%carcinoma	Somatic Mutations	Non-silent Mutations
51509/1	M	70	20	8	5
60400/2	F	30	40	8	3
218/1	M	20	50	5	2
52686/1	F	50	50	8	4
58853/3	M	0	60	20	8
60400/1	F	10	60	9	3
58163/3	M	20	70	17	6
58163/4	M	20	70	39	27
215/1	M	20	80	4	2
Total	-			**118**	**60**

Table 7

Somatic mutations detected by whole exome sequencing in *Mdr2*-KO lesions.

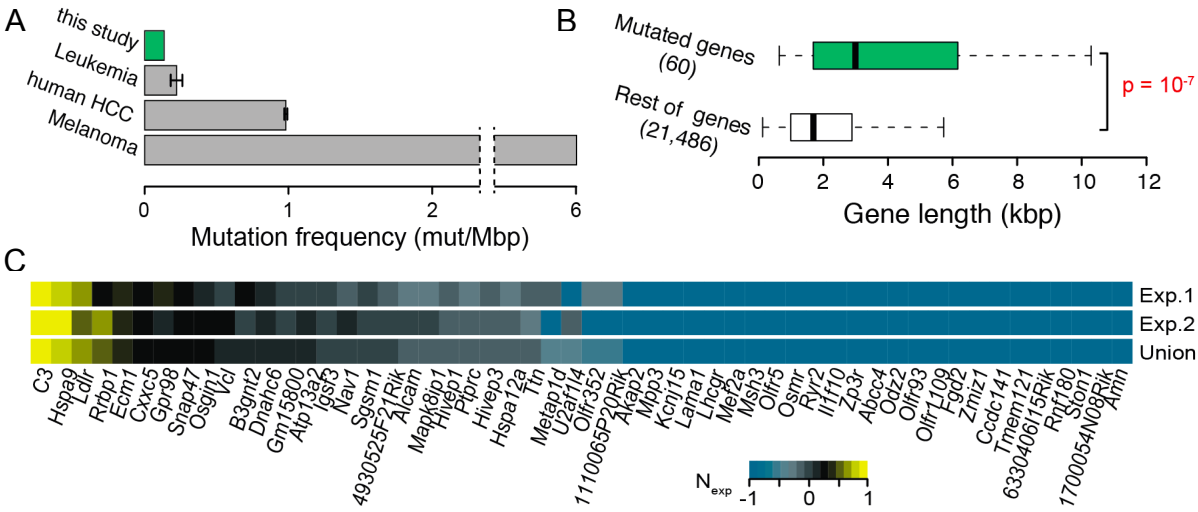


Figure 9

Spectrum of somatic mutations acquired in Mdr2-KO HCCs.

(A) Frequency of non-silent somatic mutations in the nine mouse HCCs as compared to human leukemia [85, 221-224], HCC [151, 152, 154], and melanoma [225].

(B) Gene length comparison between mutated genes and rest of mouse genes. Gene length was calculated as the total nucleotides targeted by the Agilent SureSelect XT Mouse All Exon kit. Differences in length distributions were assessed using Wilcoxon test.

(C) Expression levels of the mutated genes in normal liver [212] in the two replicates (Exp.1 and Exp.2) and the average expression over the replicates (Union). Normalized gene expression (N_{exp}) was calculated as the gene expression level over the median expression of all genes in liver. Values above and below zero indicate gene expression higher and lower than the median. -1 indicates no expression in liver.

2.2 Copy number variations assessment on *Mdr2*-KO HCCs

We then examined the incidence of copy number variations (CNVs) from the targeted sequencing data of the 9 *Mdr2*-KO exomes screened for point mutations, using a novel method based on the comparison of the normalized gene coverage between tumor and its normal counterpart (see Materials and Methods). We observed that 2,430 (~11%) of the analyzed genes displayed CNV in at least 1 sample, and that gene amplifications were the overwhelming majority (2,427) of these alterations (**Table 8**). Validation via TaqMan copy number assay of 10 predicted gene amplifications estimated 93% specificity, 85% accuracy, and 70% sensitivity of our CNV detection

method (data not shown). This evidence suggested that CNVs represent a common genetic alteration in this model.

To further assess this possibility, we performed whole genome sequencing of two HCCs extracted from two different mice (one of them belonging to the set previously analysed by whole exome sequencing), using the corresponding kidney as matched normal control. Again, we observed a predominance of amplifications over deletions, with a total of 1,074 amplified and 117 deleted genes in the two genomes (**Table 8**). The vast majority of amplified genes detected in the whole exome (85%) were also found in the whole genome, confirming the reliability of the method. Notably, the percentage of amplified genes positively correlates with the HCC content in these lesions (**Figure 10**), indicating that CNVs occur preferentially at late stages of HCC development.

Overall, CNV analysis showed that *Mdr2*-KO HCCs frequently develop CNVs, in particular gene amplifications. CNVs mostly occur at late tumorigenic stages, thus suggesting a possible driver role in tumor progression.

Lesion ID	Gender	% adenoma	%carcinoma	Targeted regions	Amplified genes	Deleted genes	Total CNVs
51509/1	M	70	20	Whole exome	59	0	59
60400/2	F	30	40		113	0	113
218/1	M	20	50		298	1	299
52686/1	F	50	50		15	2	17 *
58853/3	M	0	60		631	0	631
60400/1	F	10	60		455	0	455
58163/3	M	20	70		333	1	334
58163/4	M	20	70		625	0	625
215/1	M	20	80		562	0	562
60400/1	F	10	60	Whole genome	529	2	531
218/3	M	10	70		652	115	767
Total	-				**4,272**	**121**	**4,393**

Table 8

Copy number variations detected by whole exome and whole genome sequencing in *Mdr2*-KO lesions. In total, 2,427 amplified and 4 deleted genes were detected in the nine whole exomes, and 1,074 amplified and 117 deleted genes were found in the two whole genomes. * TaqMan copy number assay assessed high number of false negative for this sample, suggesting an overall underestimation of gene amplifications.

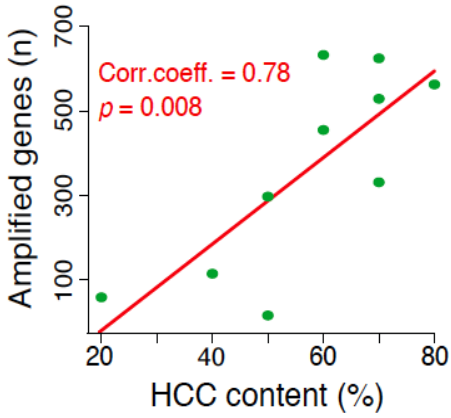


Figure 10

Number of genes amplified in 10 *Mdr2*-KO tumors as a function of the corresponding HCC content. The plot is based on the 10 mouse tumors that underwent whole exome and whole genome sequencing. Correlation was assessed using Pearson correlation test.

2.3 Analysis of recurrently altered genes in *Mdr2*-KO HCCs

In order to identify putative cancer driver genes, we focused on the 69 genes amplified in at least 50% of the sequenced HCC exomes (~2% of all altered genes, **Figure 11A**). Around 67% of them laid on chromosome 8, accounting for almost 4% of the genes on this chromosome (**Figure 11B**), implying that chromosome 8 is recurrently rearranged in *Mdr2*-KO liver cancer.

Among the top scoring genes, we found that *Map2k7* was amplified in the 4 samples (out of 9) having the highest HCC content (data not shown). *Map2k7* encodes for JNKK2/MKK7, an upstream activator of the c-Jun NH(2)-terminal kinases (JNKs) [226], which are mainly triggered by pro-inflammatory cytokines and environmental stress [227]. Since JNK pathway deregulation has already been associated with liver cancer [228, 229], *Map2k7* seemed a good candidate for further investigation. We screened 35 additional tumors from 16 distinct individuals by TaqMan copy number assay and found that *Map2k7* was amplified in 14/49 *Mdr2*-KO nodules (29%, **Table S7**), a frequency much higher than expected by chance ($p=6x10^{-04}$, binomial test). Seven of the 12 nodules composed of ≥40% HCC (58%) showed *Map2k7* amplification (**Table S7**), a fraction again significantly higher than expected by chance ($p=9x10^{-05}$, binomial test), indicating that *Map2k7* amplification occurred more frequently in lesions with high HCC content.

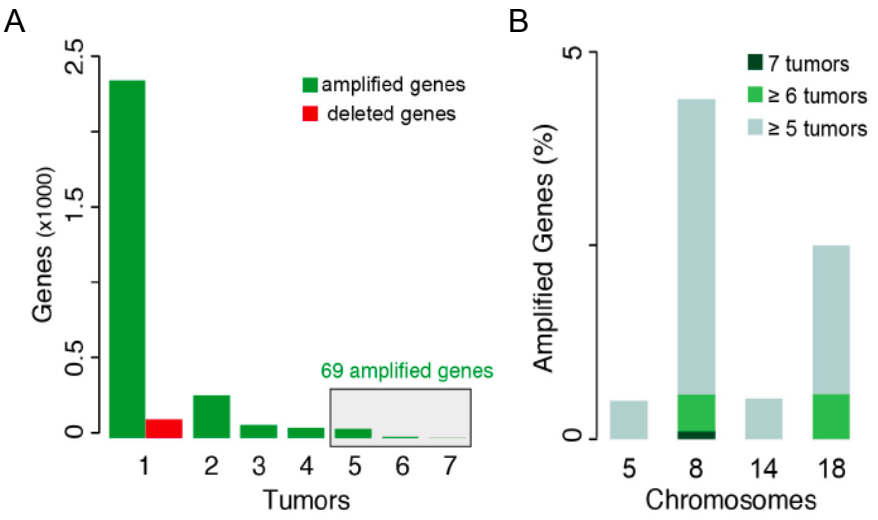


Figure 11

Recurrent gene amplifications in Mdr2-KO HCCs.

(A) All amplified (green) and deleted (red) genes in 10 tumor exomes, grouped according to the number of samples where they were altered. No gene was modified in more than seven tumors. (B) Fraction of recurrently amplified genes in each chromosome, measured as the number of amplified genes over the total number of targeted genes in that chromosome.

In order to understand whether *Map2k7* amplification impacts on its steady state mRNA levels, we measured *Map2k7* expression on the tumoral samples screened for CNVs. In particular, we compared MKK7 mRNA levels of 9 nodules that were found to bear *map2k7* amplifications with that of 7 lesions with no CNV for *map2k7*, 9 age-matched inflamed livers and 9 age-matched WT livers (**Figure 12**). MKK7 expression in the amplified samples was higher with respect to all other sets of samples. An increase in MKK7 expression was also detected in samples with normal *map2k7* copy number with respect to inflamed and WT samples. This finding suggests that an upregulation in the expression of JNK pathway components, due to either an increase in gene copy number or increased transcription, may represent a selected event in HCC development in *Mdr2*-KO mice.

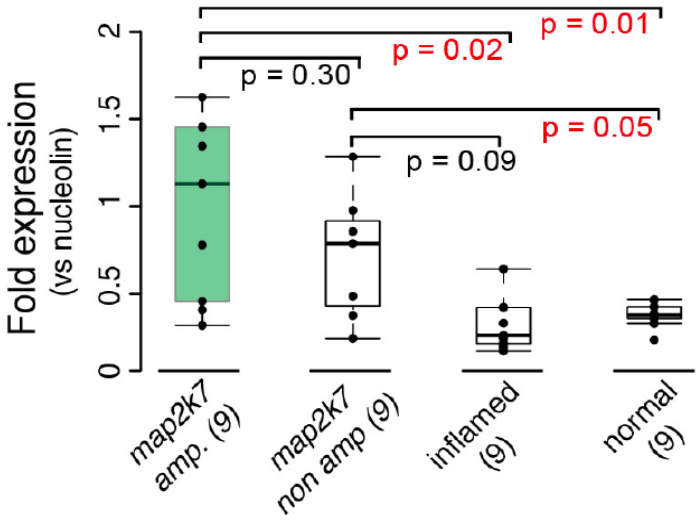


Figure 12

MKK7 expression analysis on *Mdr2*-KO HCC nodules with or without *Map2k7* amplification *vs* *Mdr2*-KO inflamed livers and *Mdr2*-WT livers. Number of screened samples are indicated in brackets. Values are expressed as percentage of fold MKK7 expression normalized on expression of nucleolin as a control gene. Differences between distributions were assessed using Wilcoxon test.

Indeed we found that all but one tumors had modifications in at least one primary JNK interactor (defined as a protein that engages direct physical interactions with the core component of the JNK pathway, **Figure 13**). We measured the distance of the proteins encoded by all altered genes in all tumors from any component of the JNK pathway in the human protein-protein interaction network and observed that they are closer to the JNK pathway than expected by chance ($p = 2x10^{-04}$, Wilcoxon test). This indicates that the alteration of other JNK pathway components in addition to *Map2k7* is a pervasive feature of *Mdr2*-KO HCC.

Collectively, these data show that gene amplifications in in *Mdr2*-KO tumors preferentially hit specific genes, most notably *Map2k7*. These gene alterations suggest that the deregulation of the JNK pathway that may have a driver role in liver tumor progression.

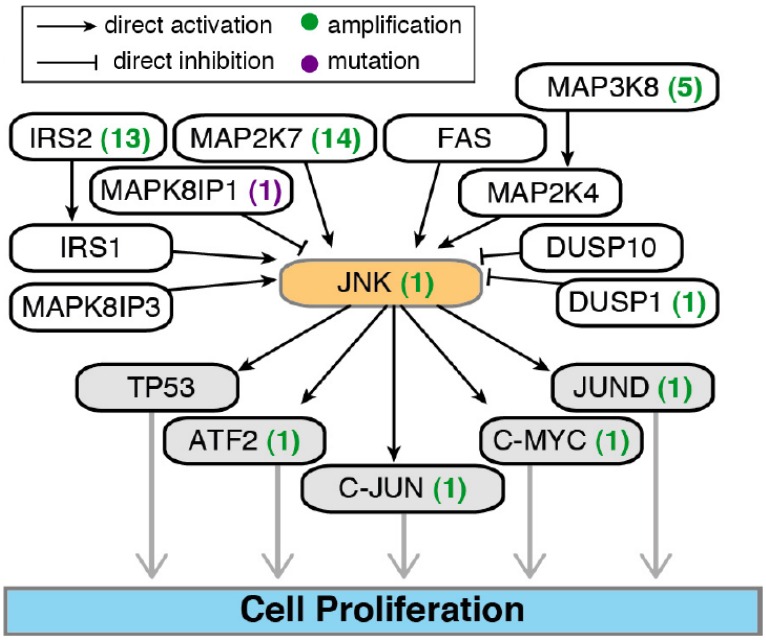


Figure 13

Network of JNK protein-protein interactions. Shown are 14 primary interactors of JNK and two secondary interactors that are recurrently mutated in *Mdr2*-KO tumors. The number of tumors where the gene was amplified (green) or mutated (purple) are reported in brackets.

2.4 Effects of inhibition of the JNK pathway on *Mdr2*-KO nodules

In order to understand whether *Map2k7* amplification was biologically relevant to HCC progression, we treated 13 to 14 months old *Mdr2*-KO mice with SP600125, a synthetic polyaromatic inhibitor of the JNK kinases [230]. After three weeks of treatment, mice were sacrificed and the tumors from the two cohorts were compared in terms of nodule number, size, histological composition, and tumor content.

Although inhibition of JNK did not have an effect on the number of HCC nodules eventually detected, it had an obvious effect on tumor size. Indeed, mice treated with SP600125 did not develop any tumors larger than 20 mm, in contrast to about 20% of mice treated with vehicle (**Figure 14A**), despite a similar proportion of nodules with *Map2k7* amplification in the two groups (**Table S8**). When considering the histological composition of the lesions, nodules bigger than 10 mm in treated mice showed significantly higher proportion of adenoma and lower proportion of carcinoma as compared to untreated mice (**Figure 14B**). No difference was detectable in the histological composition of smaller lesions. Similarly, by comparing the tumor burden per mouse in the two cohorts, treated mice showed an overall significant depletion in

HCC, while purely adenomatous nodules were over-represented (**Figure 14C**).

In summary, these results indicate JNK pathway activation as a driving requisite for *Mrd2*-KO cancer progression, and in particular for the adenoma-to-carcinoma transition in *Mrd2*-KO tumors. These data are consistent with the frequent emergence of *Map2k7* amplification in high grade HCCs (**Figure 10**).

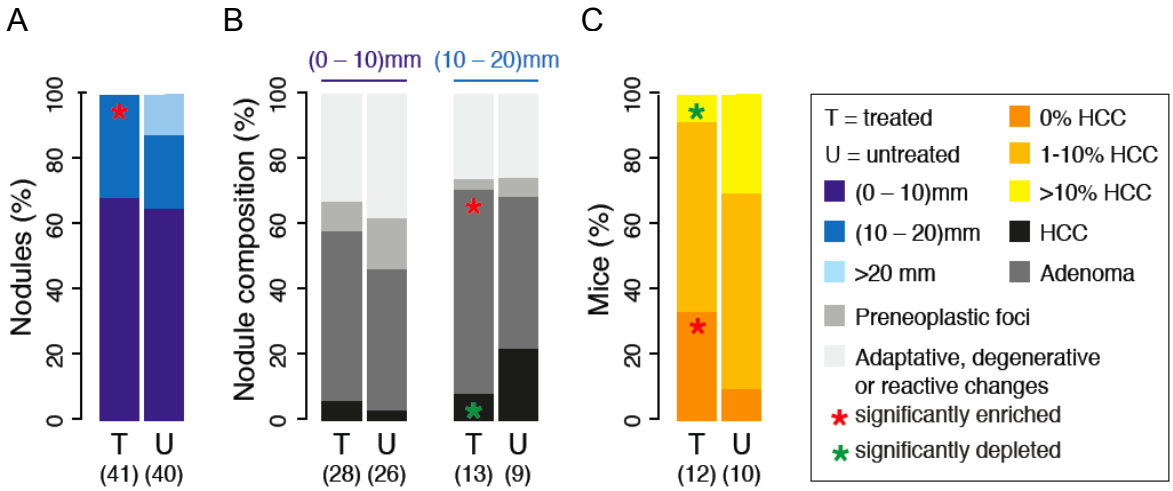


Figure 14

Effects of JNK inhibition on *Mdr2*-KO liver nodules. Analyzed nodules or mice in the two groups are reported in brackets. All differences were assessed using Fisher's exact test. See also **Table S8**.

(A) Size differences among treated and untreated nodules.

(B) Histological composition of treated and untreated nodules. Nodules in the two cohorts of mice were divided into two groups according to the size (reported on top).

2.5 Point mutation and CNV analyses on human PFIC-related HCC

The data presented insofar showed several features that distinguish HCC arising in *Mdr2*-KO mice from human HCC. First, we reported that the frequency of non-silent somatic mutations in *Mdr2*-KO HCCs is much lower than the frequency that has been measured in human liver cancer [14, 150-152, 154, 231], which is nonetheless comparable to that of other solid tumors (**Figure 9A**). Furthermore, none of the recurrently mutated genes in human HCCs [209] was found mutated in the mouse lesions that we screened. Finally, we observed that *Mdr2*-KO tumors are far more

prone to accumulate gene amplifications that deletions, which is the opposite of what found in human liver cancer [14, 150, 152, 231] (**Table 8**). These peculiar differences might be attributed to species-specific traits, implying that this mouse model of HCC does not recapitulate the genetics of the human disease. Nevertheless, the majority of patients analysed so far developed liver cancer upon exposition to virus infection or other risk factors (e.g. alcohol), which have distinct and specific impact on the acquisition of genomic alterations [209]. Therefore, as *Mdr2*-KO HCC is induced by inflammation due to cholestasis and independently from external factors, a possibility is that the observed differences might be due to different disease pathogenesis rather than to inter-species diversity.

We therefore sequenced the whole exome of HCCs from six individuals affected by Progressive Familial Intrahepatic Cholestasis (PFIC, **Table S9**). As mentioned, different PFIC forms are recognized based on the impaired liver transporter genes, and PFIC type 2, in which mutations of the *ABCB11* gene lead to bile salt export deficiency, is associated with early onset pediatric hepatocellular carcinoma [232].

In the exomes of all six human tumors, we detected a total of 44 somatic mutations, none of them being shared between the two patients, and 8 indels (**Table 9**). Of these modifications, 15 mutations and no indels were non-silent (**Table 9**). However, none of them was a known driver gene in HCC or in other human cancers [209, 210]. Moreover, as in the case of *Mdr2*-KO, most of these genes were poorly or not expressed in human normal liver [212], thus indicating that all 15 mutations were likely passenger (data not shown).

To investigate whether the massive copy number alterations observed in *Mdr2*-KO HCCs also occur in PFIC human HCC, we screened for somatic CNVs in seven PFIC-related HCCs, including all lesions screened for point mutations and one additional sample. We detected a total of 18,428 genes that underwent amplifications and 3,628 genes that underwent deletions, with an overwhelming predominance of copy number gains over losses, similar to what observed in the mouse tumors (**Table 10**). Although amplifications were generally widespread along the entire genome, chromosome 8, 19 and 20 were almost completely amplified in the majority of lesions. Indeed, amplification of chromosome 19 was confirmed in sample 23836 by fluorescence *in*

situ hybridization (FISH) with probes located on both chromosomal arms (**Figure 15**). Interestingly, human *MAP2K7* is located on chromosome 19 and it was therefore amplified in the liver of this patient. Actually, *MAP2K7* was found amplified in 5 lesions out of 7 (**Table 11**).

Altogether, these findings suggest a similar genetic landscape for purely inflammatory human and murine HCCs.

Lesion ID	Gender	% carcinoma	Somatic Mutations	Non-silent Mutations	Somatic Indels
175	M	90%	7	3	0
7860	F	90%	5	2	0
23836	M	90%	5	1	0
HB4R	F	70%	25	9	7
1790	M	60%	1	0	0
2896	F	50%	1	0	1
Total	**-**		**44**	**15**	**8**

Table 9

Somatic mutations in human PFIC-related liver cancers. None of the identified indels introduced a frameshift (*i.e.* was non-silent).

Lesion ID	Gender	% carcinoma	Screening	Amplified genes	Deleted genes	Total CNVs
175	M	90%	WES + CNV	10,688	1	10,689
7860	F	90%	WES + CNV	13,594	1,248	14,842
23836	M	90%	WES + CNV	12,450	19	12,469
HB4R	F	70%	WES + CNV	5,598	2,575	8,173
1790	M	60%	WES + CNV	8,601	0	8,601
2896	F	50%	WES + CNV	9,687	244	9,931
MB	M	40%	CNV	3,801	258	4,059
Total	**-**			**18,428**	**3,628**	**22,056**

Table 10

Copy number alterations in human PFIC-related liver cancers.

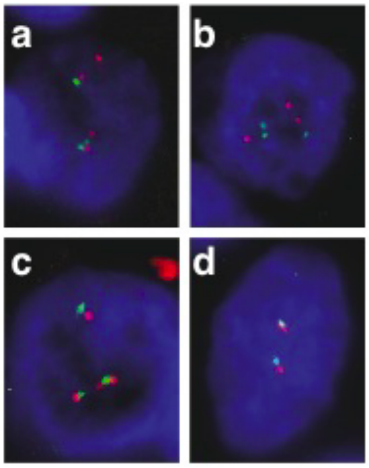


Figure 15

Amplification of chromosome 19 in patient 23836 as detected by FISH. Shown are four representative hepatocytes nuclei (blue, DAPI staining) from the paraffin-embedded tumor sample. Probes on chr19p13 (green) and on chr19q13 (orange) were used and nuclei in a-b-c show additional copies of chromosome 19, while the nucleus in d has only two.

Gene Name	Copy Numbers							Samples with amplification	Samples with deletion
	175	7860	23836	HB4R	1790	2896	MB		
MAP2K7	3	7	4	3	2	3	2	5	0
MAPK8	2	4	3	3	3	3	2	5	0

Table 11

Fraction of PFIC tumors with *Map2k7* (MMK7) and *Mapk8* (JNK) copy number alterations. For each patient the number of gene copies is indicated.

DISCUSSION

1. *Mdr2*-KO HCC is epigenetically characterized by a hepatocyte-specific inflammatory signature

The first part of this project was focused on the elucidation of epigenetic changes occurring in the context of a chronic inflammation-driven tumorigenic process. Specifically, we performed the epigenetic profiling of a panel of histone marks by ChIPseq on liver samples from *Mdr2*-KO mice, as a model of spontaneous inflammation-derived liver cancer.

We were primarily interested in the identification of TFs activated during the progression to cancer, in order to infer which molecular players are required at the chromatin level for tumor origin and development. Therefore, by applying a "reverse epigenetics" approach, we profiled H3K27Ac, a mark of active regions, to unveil TFs whose recruitment would be critical at each pathologic step (namely, chronic inflammation, adenoma, low grade HCC and high grade HCC).

We have shown that ~1,400 to 2,300 regions become hyperacetylated (with respect to the WT counterpart) in all samples of a given disease stage, and that gain in acetylation generally corresponds to a gain of H3K4me1 mark in inflamed and low grade HCC tissues.

Gene ontology analyses of these regions compared to the WTs suggested the acquisition of an inflammatory signature during the chronic inflammation stadium, possibly in response to the exposition to cytokines and other inflammatory mediators. This is in agreement with previous reports showing that hepatocytes exposed to activated macrophages *in vitro* display an up-regulation of pathways (such as the IFNγ and the TNF receptor 2 pathways) associated to inflammatory responses [233]. We showed that this inflammatory signature is retained throughout the tumorigenic process until the latest cancer stage. Interestingly, the absence of inflammatory infiltrates inside the nodular lesions implies that this signature is specific of hepatocytes and that, once triggered, its retention is independent of an inflammatory microenvironment.

The importance of the crosstalk between non-parenchymal cells and hepatocytes, both for the establishment and propagation of acute and chronic inflammation, is well known [234]. Nevertheless, to what extent hepatocytes *per se* contribute to inflammation, independently of stimuli deriving from other cell populations, is still unclear. They have been found to express mRNA for all the types of Toll-like receptors (TLRs), crucial in the recognition of pathogen-associated molecular patterns and for the anti-microbial response, as well as their adaptor molecules MyD88 and MD2 [235]. In keeping, hepatocytes have been shown to respond to lipopolysaccharide and other TLR4 ligands, although the degree of response is still matter of debate [236-238]. Moreover, stimultation of TLR3 with polyI:C caused the activation of type 1 interferon in hepatocytes [239]. Few studies on cell lines also showed the expression of specific cytokines and chemokines in stimulated hepatocytes [240-242]. Our results might point to an active involvement of hepatocytes in contributing to the inflammatory environment.

At every disease stage, we observed that DNA motifs recognized by AP1 and ETS family members were enriched in common H3K27Ac induced peaks.

The AP1 family entails a group of structurally and functionally related proteins belonging to different subfamilies, the main ones being the Jun, Fos, ATF and JDP subfamilies [217]. In common H3K27Ac-induced regions, we found enrichment of consensus sequences for all the principal AP1 subfamilies. AP1 members homo- and heterodimerize, depending on cell type-specific concentrations of each family member, and giving rise to different species of AP1 transcription factors. Upon dimerization, these TFs are able to recognize TPA (tetradecanoylphorbol acetate)-DNA responsive elements (TREs, consensus sequence 5’ -TGAG/CTCA- 3’) [218]. AP1 TFs activity is dependent on a broad range of stimuli, including cytokines, growth factors, infection, stress signaling and oncogenic factors, and is influencing key cell functions such as proliferation and apoptosis. AP1 activity is further controlled by upstream kinases, in particular by JNK [218].

The ETS family has been implicated in a number of crucial cell processes, from proliferation and differentiation to apoptosis and senescence [219]. Moreover, one of the best studied role of ETS TFs is the regulation of hematopoiesis, and some ETS

proteins are also involved in vasculogenesis, neuronal development and immunity [220, 243]. ETS is a large and variegated group of TFs, with approximately 29 members differently expressed according to cell type. The most frequently retrieved ETS matrices were associated to the SPI subfamily, and in particular to PU.1. Nevertheless, none of the SPI members are known to be expressed in liver under physiological conditions [212]. It would be interesting to measure the expression levels of PU.1, SpiB and SpiC in *Mdr2*-KO tumoral samples, in order to understand whether cancer development could cause ectoptic SPI members expression. Intriguingly, several genome-wide association studies recently found several intronic SNPs on *SPIB* locus to be a risk factor for primary biliary cirrhosis [244-246].

Moreover, further analyses will be necessary to estimate to what extent are combined AP1 and ETS motifs present in common hyperacetylated regions. In fact, one possibility is that regions acquiring H3K27Ac and becoming "open" would be bound by AP1 members only, despite the concurrent presence of an ETS motif. On the other hand we could speculate that, given the high degree of similarity between matrices associated to SPI members and other ETS subfamilies, and the high percentages of occurrence of other ETS TFs matrices reported by the FIMO analysis (**Table 6**), the ETS sites present in the common H3K27Ac induced regions are not bound by SPI TFs, but rather by other ETS proteins.

It will be however interesting to evaluate the expression levels of both AP1 and ETS family members during the progression to cancer, in order to select for relevant TF candidates for genome wide occupancy studies. For this reason, and in order to compare results with expression profiles, we are planning to perform RNAseq analyses on the same samples that have been used for ChIPseq.

By focusing on H3K27Ac mark, we chose to search for regions that become active during the disease progression. Nevertheless, important information can be obtained also from regions inactivated during the tumorigenic process. Therefore, future work will be focusing on profiling of repressive chromatin marks (i.e. H3K9me3, K3K27Ac).

Gene ontology and motif enrichment analyses failed to cluster samples according to the disease stage. This was due to the fact that the activation of inflammatory components, and the suggested involvement of AP1 TF family, was a pervasive

feature of all the *Mdr2*-KO pathologic steps starting from the development of inflammation. In order to reveal epigenetic changes specifically occurring upon cancer onset, it will be critical to perform analyses by keeping the inflamed samples (rather that the WT samples) as reference.

Finally, a crucial level of information to be added in order to have a more complete picture of epigenetic changes occurring along *Mdr2*-KO tumorigenesis would be constituted by DNA methylation. We are planning to obtain single base resolution, genome wide methylation profiles of the same samples that underwent ChIPseq analyses, in order to delineate the interplay between the two types of modifications.

2. *Mdr2*-KO mice develop late stage HCC-associated gene amplifications

The second part of the present study was aimed at assessing the impact of inflammation on the origin of genetic alterations and the consequent emergence of cancer. In particular, we screened for the presence of point mutations and copy number variations in hepatocellular carcinoma lesions from *Mdr2*-KO mice. What emerged is that CNVs, and particularly gene amplifications, is the most common mutational event in this system. Previous evidences obtained in our lab (data not shown) indicated that no genetic modifications are present at earlier pre-tumoral stages (chronic inflamed tissue, dysplastic liver). This implies that mechanisms other than genetic are the driving cause of tumor origin. By integration with the epigenetic data, it would be therefore interesting to understand whether epigenetic modifications, by regulating chromosome packing and stability, are implied in this scenario.

Results obtained by massive parallel sequencing and biological validations unveiled a key role for *Map2k7* (MKK7) in particular, and for the JNK pathway more in general, in inflammation-derived HCC. MKK7 is known to be essential for pro-inflammatory cytokines to stimulate JNK activity [226]. Furthermore, *Map2k7* is expressed in embryonic liver, where it is important to drive proliferation by upregulating the levels of the Cdc2 kinase [247]. In addition, liver-specific deletion of *Mapk14*, a negative regulator of *Map2k7*, leads to JNK hyperactivation and higher tumor development

[248]. MKK7 oncogenic potential is nevertheless still controversial, as it was recently described to play a tumor-suppressive action in murine lung and mammary cancer [249].

However, the implication of JNK in liver cancer is well known. Lack of c-Jun dramatically reduces liver carcinogenesis in DEN-induced mouse HCC model [250], and *Jnk*$^{-/-}$ mice exhibit reduced liver tumors [251]. Proliferation of human and murine HCC was shown to require JNK1-dependent p21 downregulation [252]. Importantly, 70% of human HCC tissues have been found to be positive for phospho-JNK immunostaining [253]. Chronic activation of TGF-β activated kinase 1 (TAK1) and JNK due to disruption of cylindromatosis tumor suppressor (*Cyld*) gene was recently described to cause massive hepatocyte death accompanied by compensatory hepatocyte proliferation, which leads to development of HCC [228]. Despite all these indications underlining the importance of JNK in liver cancer, whether JNK pathway activation is occurring in hepatocytes rather than in other non-parenchymal liver cells is still matter of debate. Das and colleagues suggested that JNK seems to promote HCC formation in non-parenchymal cells by increasing the production of pro-inflammatory cytokines, while playing an opposite role in hepatocytes, where combined deficiency of *Jnk1* and *2* is linked to increased tumor size in DEN- treated mice [229].

Nevertheless, our work contributes to demonstrating an alteration of JNK pathway (and of MKK7 in particular) specifically into hepatocytes, which does not exclude a parallel misregulation inside other types of liver cells. Furthermore, we show that deregulation of JNK pathway in HCC is due to a genetic alteration. In particular, we saw that JNK amplifications generally occur at late stage of HCC development, therefore the deregulation of this pathway is likely to favour cancer progression rather than tumor initiation. Amplification of *Map2k7,* as well as recurrent alteration of JNK pathway components, has not been reported in any WES or WGS study on human HCC [14, 150-152, 154, 231]. Interestingly, and relevant to the first part of this study, JNK is one of the main upstream activators of AP1 family components.

In order to understand whether the pattern of modifications depicted in this work is directly due to the inflammatory context and it is not simply specific to the mouse

model (which, despite approximating human HCC pathogenesis, surely constitutes a different background), we were interested in testing whether the same genetic lesions were occurring in human HCC patients. In an attempt to chose an exclusively inflammation-driven cancer, and given the importance of the disease etiology, we validated our findings on samples deriving from PFIC patients. Importantly, we found that these patients display a genetic alteration pattern very similar to the one envisaged for *Mdr2*-KO mice, with no candidate driver mutations, very low mutation rates, and high CNV frequency with predominance of CN amplifications. Taken together, these evidences demonstrate that *Mdr2*-KO mice closely recapitulate the genomic lesions occurring in PFIC patients, and confirm that HCC deriving from a purely inflammatory background has a peculiar genomic landscape that markedly differs from viral HCC.

In conclusion, this part shows that inflammation *per se* promotes the generation of a distinctive genomic signature in the liver, which is sharply different from those originating from virus- and chemical-induced inflammation. In particular, inflammation-derived HCC does not display genomic alterations in the earliest tumorigenic steps, and is accompanied by recurrent gene amplifications (particularly on *Map2k7* and other JNK pathway members) during tumor progression.

3. Working model and future perspectives

With this work, we aimed at detecting mutations and epigenetic modifications induced by chronic inflammation at each stage of tumor origin and development. We have shown that, during a chronic inflammation-driven tumorigenic process, an activation of JNK-AP1 components is triggered early upon tissue inflammation, and is persistence throughout disease progression. Upon formation of nodular lesions, despite segregation from the surrounding inflammatory *milieu*, an epigenetic inflammatory signature acquired by hepatocytes is maintained, and at the latest stages of tumor development this correlates to genetic alteration events that lead to JNK pathway amplification (**Figure 16**).

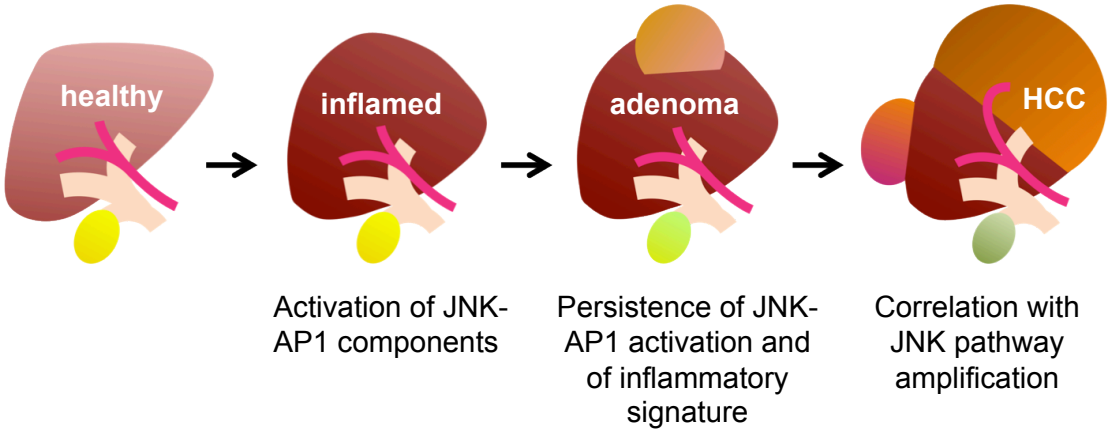


Figure 16

Working model.

Future work will be necessary to corroborate this working model. In particular, in order to definitely clarify whether the JNK pathway is functional to tumorigenesis, an interesting possibility would be to inhibit JNK or Jun in *Mdr2*-KO mice specifically in hepatocytes. It would be possible to cross *Mdr2*-KO mice with c-jun$^{f/f}$ mice [254], in order to genetically knock out the JNK pathway. Nevertheless, this could influence the peculiar genetic landscape that drives tumorigenesis in the *Mdr2*-KO model. Another possibility would be to abrogate JNK pathway *via* RNA interference (RNAi), by adenoviral infection of shRNA (short hairpin RNA) or by hydrodynamic pump injection of naked siRNA (small interfering RNA) [255, 256].

Nevertheless, this project primarily represents an effort toward producing an integrated analysis of the nuclear changes occurring in a murine model of cancer, by the generation of genome-wide, multi-stage datasets for multiple levels of complexity (genomic, epigenomic and, in future, transcriptomic). The long-term goal in this process, once more exhaustive information from each level of complexity will be achieved, will be to funnel these variegated datasets into an integrated systems-level model, in order to determine the temporal relationship and interplay of modifications in different compartments. This will help in shedding light on the complex system of interactions between genomic and epigenomic machineries and how alterations in this crosstalk can lead to disease.

SUPPLEMENTARY FIGURES AND TABLES

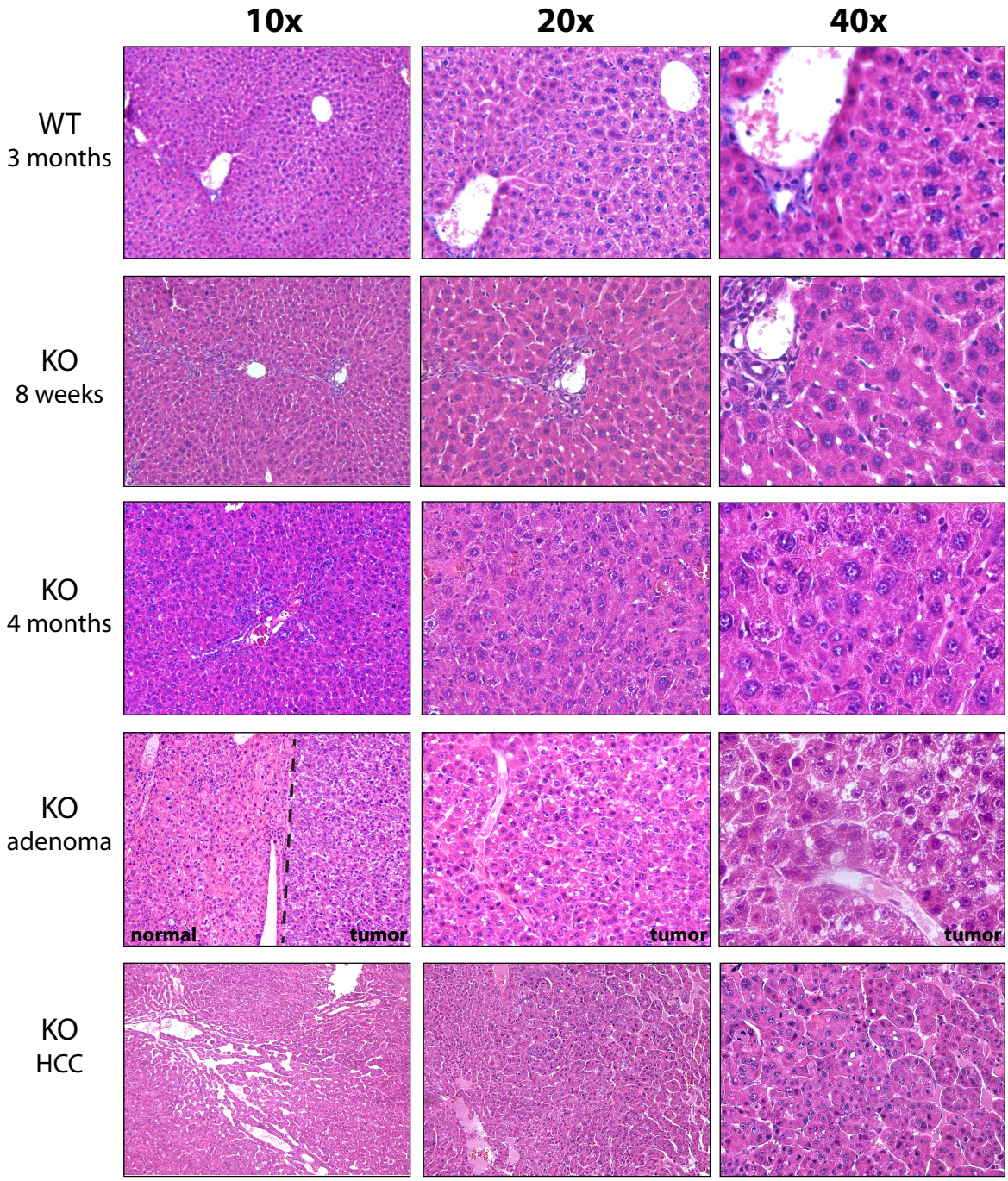


Figure S 1

Hematoxylin/eosin histological analysis of livers from *Mdr2*-KO and *Mdr2*-WT mice. At 8 weeks ductular proliferation is visible despite a general architecture conservation. Hyperproliferation and severe architectural dysplasia are observed at 4 months, which leads to development of incapsulated adenoma between 9 and 12 months and, subsequently, to end-stage tissue disruption and HCC.

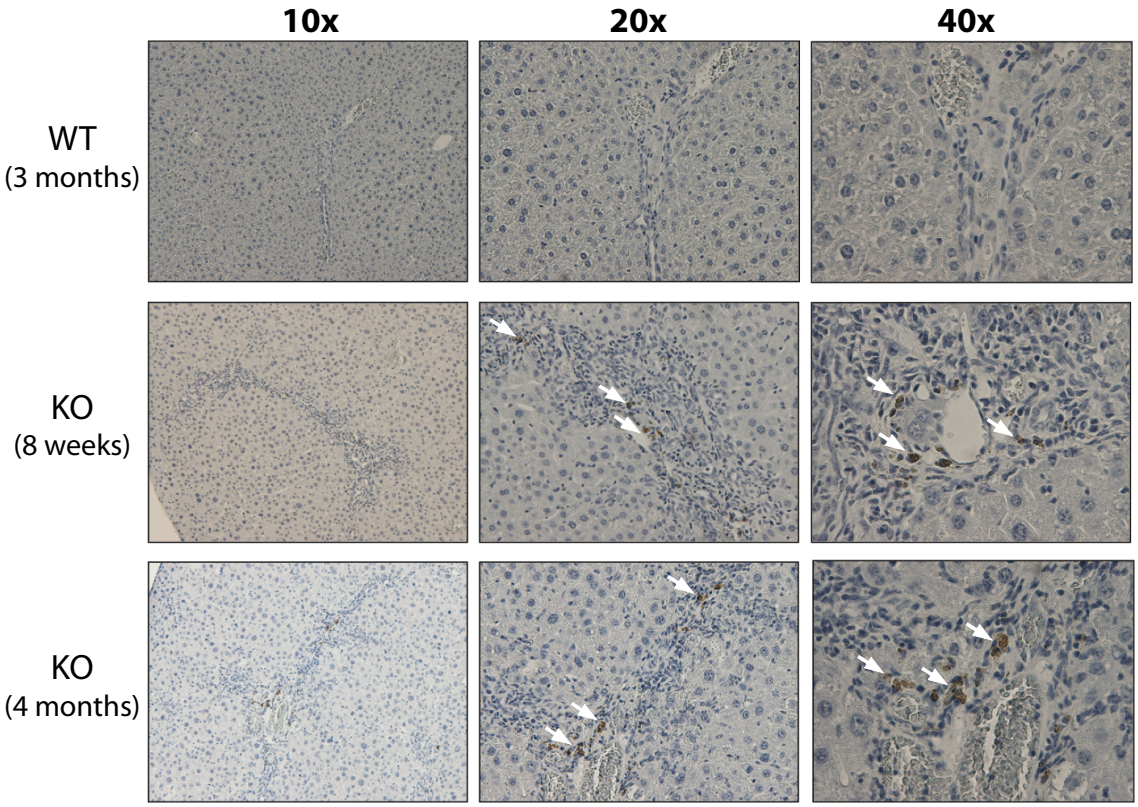


Figure S 2

Immunohistochemistry on livers from *Mdr2*-KO and *Mdr2*-WT mice (8 weeks and 4 months old, same individuals as in **Figure S1**). Liver slices have been stained with a biotinylated anti-CD3 antibody to visualize inflammatory infiltrate in the periportal parenchymal tissue of *Mdr2*-KO mice.

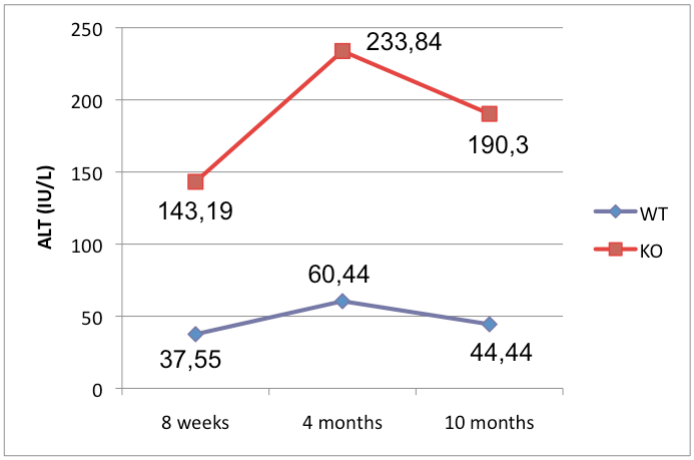


Figure S 3

Measurement of alanine aminotransferase (ALT) plasma levels on *Mdr2*-KO and *Mdr2*-WT mice (8 weeks, 4 months and 10 months old).

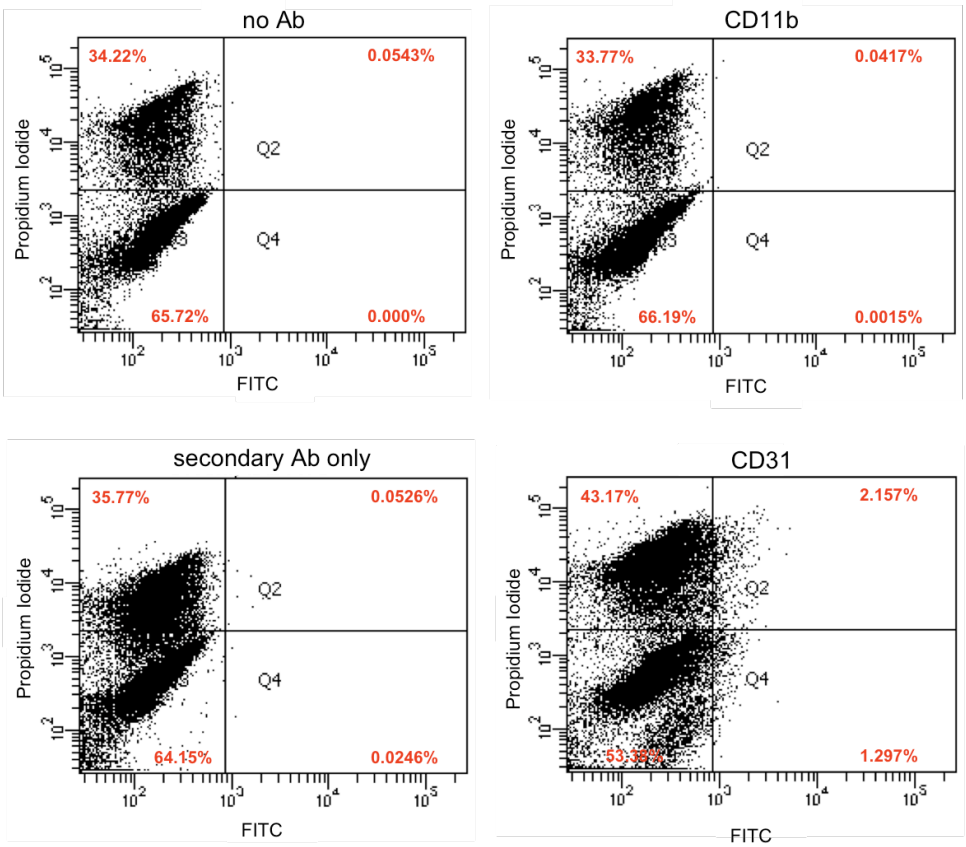
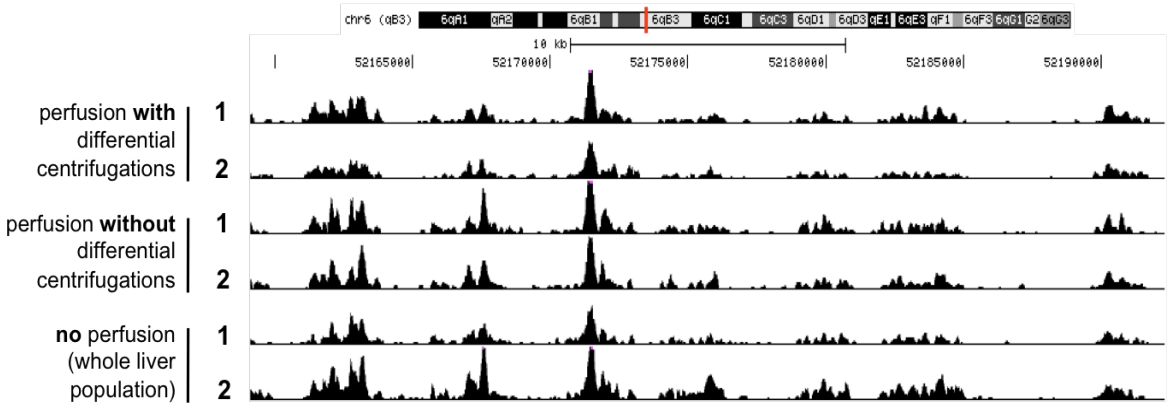


Figure S 4

FACS analysis to evaluate purity and viability of a hepatocyte preparation obtained by liver perfusion. Due to the lack of a reliable membrane marker for hepatocytes, cells have been stained for CD11b and CD31, to mark Kupffer and endothelial cells, respectively (the two most represented populations in liver after hepatocytes). As reported, cell contaminants in our preparations is very low (0,0015% for Kupffer and 1,297% for endothelial cells). Viability is evaluated with propidium iodide, and the percentage of dead cells (between 34,22% and 43,17%) is similar to the percentage observed by Trypan blue inclusion.



		perfusion with centrifugations		perfusion without centrifugations		no perfusion	
		1	**2**	**1**	**2**	**1**	**2**
perfusion with centrifugations	**1**	1	0.76	0.74	0.74	0.72	0.68
	2		1	0.72	0.72	0.72	0.63
perfusion without centrifugations	**1**			1	0.75	0.73	0.71
	2				1	0.72	0.71
no perfusion	**1**					1	0.67
	2						1

Figure S 5

Preliminary H3K4me1 ChIPseq to evaluate possible differences due to the hepatocytes or tissue preparation protocol. In order to exclude data variability due to differences introduced by the extraction of hepatocytes by liver perfusion with respect to the direct liver excision and homogenization, livers from two 8 weeks old *Mdr2*-WT animals have been processed with each protocol. A third condition has been introduced, by eliminating the final centrifugation step from the perfusion protocol, which allows to separate hepatocytes from non-parenchymal cells, but at the same time could modify chromatin (as it is performed prior to cell fixation). A representative snapshot of the ChIPseq tracks is shown (on top), along with the tracks correlation values (at the bottom). The low differences observed among different procedures has been confirmed also by CAGE transcriptome analysis (K. Hashimoto, P. Carninci, personal communication).

Stage	Sample	N. total reads	N. mappable reads	N. reads after MACS filtering
H3K27Ac				
WT	1	27,770,349	23,259,196	21,411,218
	2	28,874,387	24,190,080	23,048,350
	3	18,323,808	12,369,490	9,931,612
	4	22,793,087	19,972,486	18,723,015
	Pool	28,606,217	24,286,097	24,134,260
Inflamed	1	26,967,372	24,201,794	23,103,068
	2	39,897,199	33,984,279	31,069,231
	3	34,343,371	29,210,045	28,566,775
	4	23,233,563	19,913,343	19,140,602
Adenoma	1	40,739,350	37,053,973	37,053,973
	2	38,107,393	33,730,908	32,491,297
	3	33,684,721	30,579,301	29,962,831
	4	24,536,935	22,106,399	21,473,683
Low grade HCC	1	43,696,757	39,030,007	32,091,423
	2	34,657,078	30,312,762	29,049,282
	3	25,785,873	22,117,795	21,616,900
	4	26,779,208	19,078,268	15,015,739
	5	22,427,820	19,787,898	19,114,527
High grade HCC	1	40,592,567	36,349,089	34,719,289
	2	7,953,798	7,066,028	6,299,431
	3	38,810,413	34,587,358	33,461,884
	Pool	17,395,259	15,364,786	15,192,535
H3K4me1				
WT	1	70,979,664	56,934,985	53,380,695
	2	35,471,038	28,557,650	28,054,184
	3	27,263,257	17,153,262	11,619,052
	4	61,380,255	46,655,207	32,478,331
	Pool	69,565,992	52,255,783	51,893,368
Inflamed	1	36,411,874	31,607,891	30,395,234
	2	46,346,414	37,197,040	35,287,692
	3	89,779,177	70,012,477	66,133,061
	4	53,262,889	44,190,787	42,925,259
Low grade HCC	1	53,875,025	47,387,026	42,386,451
	2	26,166,831	21,962,891	21,501,428
	3	67,355,899	56,348,222	55,296,406
	4	40,264,844	20,789,839	13,703,868
	5	38,729,729	33,887,751	31,864,028
H3K4me3				
WT	1	18,278,734	8,965,391	3,451,238
	2	39,155,420	27,431,430	19,979,750
	3	14,042,479	7,629,898	3,779,985
	4	32,712,484	23,799,122	14,054,666
	Pool	27,464,787	19,650,752	18,002,143
Inflamed	1	40,343,074	27,026,764	21,850,997
	2	34,572,032	21,386,508	18,736,487
	3	33,379,450	17,409,400	16,076,518
	4	32,203,587	18,583,977	17,250,866
Low grade HCC	1	29,491,423	20,107,086	15,271,835
	2	21,278,833	12,571,038	11,658,429
	3	21,568,234	13,821,282	12,477,976
	4	20,525,425	13,758,851	10,584,222
	5	25,251,303	15,811,058	13,736,223

Table S 1

Sequencing throughput for ChIPseqs on *Mdr2*-WT and *Mdr2*-KO liver samples. Indicated are the total number of raw reads, the number of reads aligned to the genome, and the number of reads retained after MACS filtering.

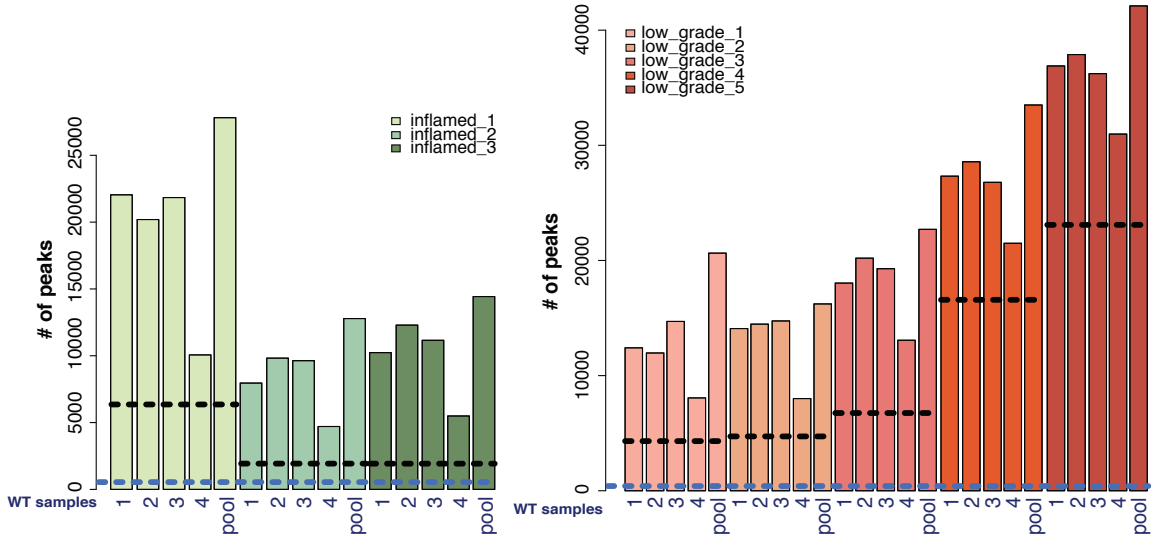


Figure S 6

Number of peaks for *Mdr2*-KO H3K4me1 ChIPseq samples. Numbers of the WT samples (in blue) are reported at the bottom of each bar. Dashed black lines indicate the number of common peaks of each sample when compared to each of the WT samples, while dashed blue lines indicate common peaks among all samples compared to all WTs.

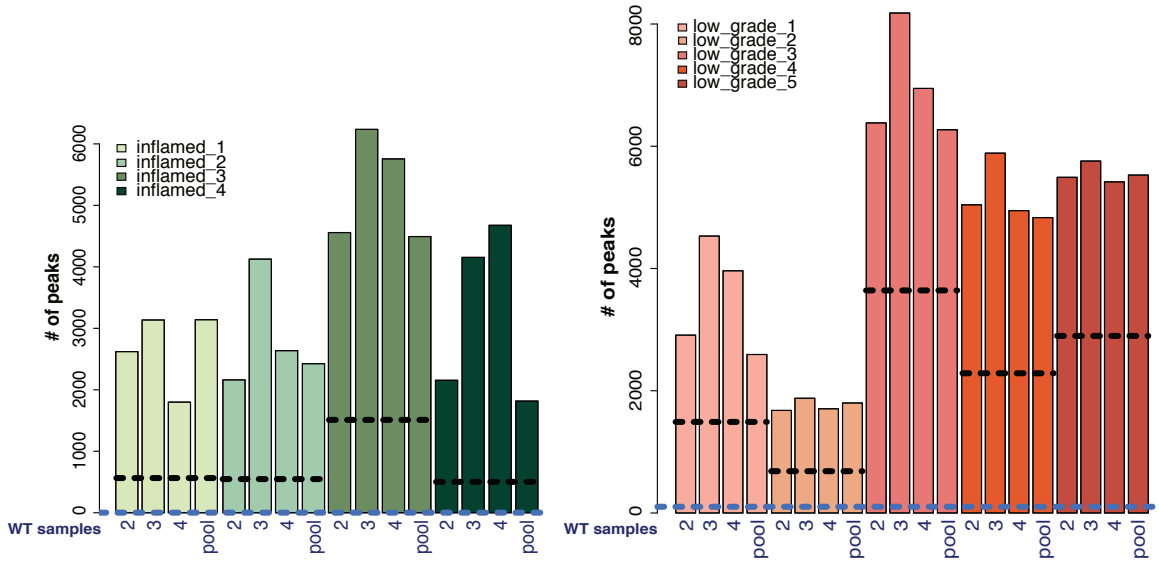


Figure S 7

Number of peaks for *Mdr2*-KO H3K4me3 ChIPseq samples. Numbers of the WT samples (in blue) are reported at the bottom of each bar. Dashed black lines indicate the number of common peaks of each sample when compared to each of the WT samples, while dashed blue lines indicate common peaks among all samples compared to all WTs.

# Term Name	Binom Rank	Binom Raw P-Value	Binom FDR Q-Value
immune system process	1	5.25E-47	4.50E-43
regulation of immune system process	2	5.16E-45	2.21E-41
positive regulation of immune system process	3	5.36E-32	1.53E-28
response to wounding	5	5.10E-29	8.75E-26
regulation of immune response	8	7.33E-27	7.87E-24
cell activation	13	2.63E-22	1.74E-19
response to biotic stimulus	14	4.08E-22	2.50E-19
response to other organism	16	1.27E-21	6.80E-19
regulation of response to stress	17	1.36E-21	6.89E-19
defense response	18	7.81E-21	3.72E-18
leukocyte chemotaxis	19	1.62E-20	7.33E-18
phagocytosis	20	1.79E-20	7.70E-18
endocytosis	21	2.38E-20	9.72E-18
immune response	22	2.56E-20	9.99E-18
regulation of cell activation	24	3.39E-20	1.21E-17
leukocyte activation	25	4.73E-20	1.62E-17
leukocyte migration	26	6.95E-20	2.29E-17
wound healing	27	1.81E-19	5.76E-17
regulation of myeloid leukocyte mediated immunity	28	2.48E-19	7.59E-17
cellular membrane organization	29	3.93E-19	1.16E-16

Table S 2

Top 20 "Biological Processes" terms of GREAT analysis on H3K27Ac common ChIPseq peaks from inflamed samples.

# Term Name	Binom Rank	Binom Raw P-Value	Binom FDR Q-Value
regulation of immune system process	2	1.62E-28	6.97E-25
positive regulation of programmed cell death	8	9.26E-22	9.94E-19
positive regulation of cell death	9	6.66E-21	6.35E-18
positive regulation of apoptosis	10	6.75E-21	5.79E-18
regulation of cell activation	17	6.50E-18	3.28E-15
induction of programmed cell death	18	7.24E-18	3.45E-15
regulation of T cell activation	19	1.34E-17	6.04E-15
regulation of leukocyte activation	24	4.42E-17	1.58E-14
induction of apoptosis	26	1.27E-16	4.18E-14
regulation of lymphocyte activation	29	2.69E-16	7.96E-14
actin filament-based process	31	8.51E-15	2.36E-12
response to drug	32	8.84E-15	2.37E-12
negative regulation of immune system process	34	1.17E-14	2.95E-12
regulation of T cell proliferation	41	1.36E-13	2.85E-11
regulation of B cell proliferation	45	1.71E-13	3.26E-11
actin cytoskeleton organization	46	2.05E-13	3.82E-11
regulation of B cell activation	58	1.77E-12	2.62E-10
regulation of immune response	59	1.78E-12	2.59E-10
negative regulation of cell migration	63	2.20E-12	3.00E-10
negative regulation of cell activation	65	2.83E-12	3.74E-10

Table S 3

Top 20 "Biological Processes" terms of GREAT analysis on H3K27Ac common ChIPseq peaks from adenoma samples.

# Term Name	Binom Rank	Binom Raw P-Value	Binom FDR Q-Value
regulation of immune system process	1	8.31E-36	7.13E-32
response to wounding	3	8.48E-26	2.43E-22
wound healing	4	5.43E-25	1.16E-21
actin filament-based process	8	2.40E-20	2.57E-17
positive regulation of angiogenesis	9	2.56E-20	2.44E-17
actin cytoskeleton organization	12	8.25E-20	5.90E-17
negative regulation of immune system process	17	4.77E-18	2.41E-15
regulation of cellular component movement	18	8.34E-18	3.98E-15
regulation of cell migration	19	1.19E-17	5.39E-15
regulation of cell motility	20	1.38E-17	5.94E-15
positive regulation of cell differentiation	21	1.54E-17	6.31E-15
leukocyte chemotaxis	23	2.14E-17	7.99E-15
regulation of angiogenesis	25	2.59E-17	8.90E-15
positive regulation of apoptosis	26	5.91E-17	1.95E-14
regulation of cell activation	27	7.29E-17	2.32E-14
cytoskeleton organization	28	7.72E-17	2.37E-14
positive regulation of programmed cell death	30	8.51E-17	2.44E-14
regulation of leukocyte activation	32	1.48E-16	3.98E-14
cell chemotaxis	33	2.20E-16	5.73E-14
regulation of lymphocyte activation	34	2.74E-16	6.91E-14

Table S 4

Top 20 "Biological Processes" terms of GREAT analysis on H3K27Ac common ChIPseq peaks from low grade HCC samples.

# Term Name	Binom Rank	Binom Raw P-Value	Binom FDR Q-Value
regulation of immune system process	2	3.09E-30	1.33E-26
response to wounding	9	4.85E-25	4.62E-22
immune response	13	4.44E-22	2.93E-19
actin filament-based process	16	3.20E-20	1.72E-17
actin cytoskeleton organization	18	2.51E-19	1.20E-16
phagocytosis	22	5.29E-18	2.07E-15
regulation of cell activation	25	6.59E-17	2.26E-14
induction of programmed cell death	30	8.26E-16	2.36E-13
regulation of myeloid leukocyte mediated immunity	31	9.39E-16	2.60E-13
induction of apoptosis	35	1.69E-15	4.14E-13
regulation of immune response	38	3.00E-15	6.78E-13
regulation of leukocyte activation	40	5.36E-15	1.15E-12
tissue remodeling	43	5.96E-15	1.19E-12
endocytosis	44	6.66E-15	1.30E-12
regulation of leukocyte degranulation	45	7.91E-15	1.51E-12
leukocyte migration	47	9.48E-15	1.73E-12
regulation of cytokine production	50	1.30E-14	2.24E-12
positive regulation of angiogenesis	55	3.39E-14	5.30E-12
regulation of angiogenesis	56	7.58E-14	1.16E-11
inflammatory response	57	8.73E-14	1.32E-11

Table S 5

Top 20 "Biological Processes" terms of GREAT analysis on H3K27Ac common ChIPseq peaks from high grade HCC samples.

Lesion ID	Sequenced regions	Sequenced Mbps	Samples per lane	Sequencing setting	Raw Gbp (mean)	Mean coverage
51509/1	Whole exome	50.4	2	101 PE	22.4	126
60400/2						142
218/1						138
52686/1						154
58853/3						146
60400/1						124
58163/3						155
58163/4						105
215/1						159
218/3	Whole genome	2,600	1	101 PE	43.5	11
60400/1						13

Table S 6

Sequencing settings, throughput and coverage for genomic analyses in *Mdr2*-KO HCCs. Details on the Illumina sequencing are provided, such as the number of samples per lane, the setting (PE = paired end sequencing), the mean raw sequenced gigabases (Gbps), and the mean coverage of aligned reads in each sample.

Lesion ID	Gender	Age (months)	Size (mm)	%carcinoma	TaqMan validation	Screening
55484/2	F	10	1	N.A.	WT	-
55484/5	F	10	1	N.A.	WT	-
54913/3	F	10	2	N.A.	Amplified	-
54913/4	F	10	1	N.A.	WT	-
54913/6	F	10	1	N.A.	WT	-
54913/10	F	10	1	N.A.	Amplified	-
54913/8	F	10	5	N.A.	Amplified	-
55481/1	F	10	4	N.A.	WT	-
55481/4	F	10	3	N.A.	WT	-
55481/2	F	10	3	N.A.	WT	-
55481/8	F	10	1	N.A.	WT	-
55481/9	F	10	3	N.A.	WT	-
55481/10	F	10	3	N.A.	Amplified	-
58859/1	F	12	6	N.A.	WT	-
52682/1	M	11	4	N.A.	WT	-
52682/2	M	11	5	N.A.	WT	-
58855/1	M	12	5	N.A.	WT	-
58855/2	M	12	6	N.A.	WT	-
55484/1	F	10	3	0	WT	-
51505/1	M	14	7	0	WT	-
60123/1	M	14	6	0	Amplified	-
52687/1	F	15	13	0	WT	-
51509/5	M	16	11	0	Amplified	-
60400/3	F	13	5	10	WT	-
60400/4	F	13	5	10	WT	-
58853/1	M	15	15	10	WT	-
58853/2	M	15	40	10	WT	-
58853/4	M	15	10	10	WT	-
58163/1	M	15	40	10	WT	-
58163/2	M	15	13	10	WT	-
52686/1	F	15	10	20	WT	-
52687/2	F	15	9	20	WT	-
51509/1	M	16	15	20	WT	-
51509/2	M	16	11	20	WT	-
51509/3	M	16	8	20	WT	-
51509/4	M	16	11	20	WT	-
55484/4	F	10	30	30	Amplified	-
60400/2	F	13	14	40	Amplified	WES
55484/3	F	10	3	50	WT	-
218/1	M	15	10	50	WT	-
52686/2	F	15	7	50	Amplified	WES
60400/1	F	13	9	60	Amplified	WES;-WGS
58853/3	M	15	17	60	WT	-
218/3	M	15	10	70	Amplified	WGS
58163/3	M	15	30	70	Amplified	WES
58163/4	M	15	30	70	WT	-
51505/2	M	14	N.A.	80	Amplified	-
215/1	M	14	18	80	Amplified	WES
218/2	M	15	30	90	WT	-

Table S 7

Results of the *Map2k7* TaqMan copy number assay on 49 *Mdr2*-KO nodules. For some nodules, histology could not be performed (N.A. = not applicable). WT indicates a normal copy number for *Map2k7*. Nodules that where previously used for WES or WGS are included and indicated.

Lesion ID	Age (months)	Group	Size (mm)	Nodule histological composition (%)					TaqMan validation
				Normal	Degenerative changes	Preneoplastic foci	Adenoma	HCC	
79568/1	14	U	11.22	0	50	0	50	0	Amplified
79568/2	14	U	NA	0	80	20	0	0	-
79568/4	14	U	2	0	40	20	40	0	-
79568/5	14	U	9.37	0	40	10	50	0	-
81210/1	14	U	22.12	0	30	0	70	0	-
81210/10	14	U	6.8	0	30	0	60	10	-
81210/11	14	U	9	0	20	10	60	10	-
81210/12	14	U	NA	0	20	20	60	0	-
81210/2	14	U	NA	0	30	30	40	0	-
81210/3	14	U	NA	0	40	40	20	0	-
81210/4	14	U	NA	0	30	0	70	0	-
81210/5	14	U	10	0	50	10	40	0	-
81210/7	14	U	NA	0	70	0	30	0	-
81210/8	14	U	4	0	70	30	0	0	-
81210/9	14	U	7.3	0	0	0	100	0	-
81217/1	14	U	21.87	0	0	0	70	30	-
81217/2	14	U	NA	0	70	0	30	0	-
81217/3	14	U	NA	0	10	10	80	0	-
81217/4	14	U	NA	0	30	10	60	0	-
81217/5	14	U	11.8	0	10	0	90	0	WT
81217/6	14	U	13.6	0	40	1	50	0	WT
81215/1	14	U	6.77	0	0	0	70	30	-
81215/2	14	U	2.9	0	70	10	20	0	-
81215/3	14	U	4.26	0	70	10	20	0	-
81215/4	14	U	4.7	0	10	10	80	0	-
81215/5	14	U	5.5	0	40	30	30	0	-
82367/1	14	U	6.9	0	50	10	30	10	-
82367/2	14	U	3.8	0	50	50	0	0	-
82367/3	14	U	NA	0	40	20	30	10	-
82367/5	14	U	2	0	40	40	20	0	-
82367/7	14	U	3.8	0	40	40	20	0	-
82367/8	14	U	NA	0	40	20	40	0	-
82367/9	14	U	NA	0	30	20	50	0	-
82367/LL	14	U	25.48	0	30	30	30	10	-
81223/1	14	U	10.13	0	20	0	70	10	Amplified
81223/4	14	U	7.58	0	40	10	40	10	-
81223/5	14	U	8.79	0	40	10	40	10	-
81223/6	14	U	NA	0	10	10	80	0	-
81223/7	14	U	5	0	10	10	80	0	-
83102/1	13	U	10.6	0	30	20	50	0	WT
83102/2	13	U	NA	0	40	20	40	0	-
83102/3	13	U	NA	0	60	20	20	0	-
83102/5	13	U	3.9	0	10	10	80	0	-
83102/6	13	U	2.8	0	60	10	30	0	-
83102/LL	13	U	37	0	10	0	40	50	-
83112/1	13	U	11.3	0	0	0	40	60	Amplified
83112/2	13	U	NA	0	70	20	10	0	-
83112/3	13	U	2	0	70	30	0	0	-
83112/4	13	U	8.3	0	70	10	20	0	-
83112/5	13	U	NA	0	40	0	60	0	-
83112/6	13	U	18.18	0	20	0	60	20	Amplified
87184/1	13	U	10.99	0	30	10	10	50	WT
87184/RL	13	U	NA	0	40	10	40	10	-
87183/1	13	U	20.1	0	40	0	50	10	-
87183/2	13	U	10	0	10	0	80	10	-
87183/3	13	U	6.3	0	20	10	70	0	-
87183/5	13	U	11	0	30	10	0	60	WT
87186/1	13	U	2	N.A.	N.A.	N.A.	N.A.	N.A.	-
88000/1	13	T	18.21	0	20	0	70	10	Amplified
81201/1	14	T	17.16	0	40	0	50	10	WT
81201/5	14	T	8.4	N.A.	N.A.	N.A.	N.A.	N.A.	-
81203/1	14	T	15.5	0	30	0	70	0	WT
81203/3	14	T	8	0	90	10	0	0	-
81203/4	14	T	2.5	0	30	0	70	0	-
81203/5	14	T	4.3	0	30	0	50	20	-
81203/7	14	T	NA	0	40	0	60	0	-
81221/1	14	T	10.53	0	20	0	80	0	WT
81221/2	14	T	3	0	40	10	50	0	-
81221/3	14	T	9	0	20	20	60	0	-
81221/4	14	T	8.2	0	20	10	70	0	-
81221/5	14	T	NA	0	50	30	20	0	-
81221/6	14	T	NA	0	30	10	60	0	-
81221/7	14	T	NA	0	0	0	100	0	-

81221/8	14	T	NA	0	100	0	0	0	-
88002/1	13	T	18.35	0	40	0	60	0	WT
88002/3	13	T	6	0	40	0	60	0	-
88002/5	13	T	3.5	0	60	0	40	0	-
82368/1	14	T	8.44	0	0	0	100	0	-
82368/2	14	T	NA	0	20	40	40	0	-
82368/3	14	T	NA	0	60	10	30	0	-
82368/4	14	T	NA	0	40	10	40	10	-
83099/1	13	T	19.4	0	0	0	80	20	WT
83099/2	13	T	7.7	0	0	0	80	20	-
83099/10	13	T	NA	0	70	20	10	0	-
83099/11	13	T	4.8	0	40	20	20	20	-
83099/3	13	T	16.5	0	10	0	90	0	WT
83099/4	13	T	16.5	0	10	0	90	0	Insufficient DNA
83099/6	13	T	7.2	0	20	10	40	30	-
83099/7	13	T	8	0	10	20	60	10	-
81214/1	14	T	17.49	0	20	0	80	0	WT
81214/2	14	T	2	0	60	10	30	0	-
81214/3	14	T	10.74	0	40	10	50	0	WT
81214/4	14	T	5.9	0	20	10	70	0	-
81214/5	14	T	5	0	20	40	40	0	-
81214/8	14	T	2	0	20	0	70	10	-
81214/9	14	T	2	0	20	0	70	10	-
81214/11	14	T	2	0	20	0	70	10	-
81214/12	14	T	2	0	20	0	70	10	-
81214/7	14	T	14.58	0	20	10	20	50	Wt
83111/1	13	T	NA	0	100	0	0	0	-
83111/2	13	T	NA	0	20	40	40	0	-
87187/1	13	T	18.84	0	50	10	30	10	Amplified
87187/2	13	T	6.5	0	30	10	60	0	-
87187/3	13	T	5.9	0	10	0	70	20	-
87187/6	13	T	NA	0	80	20	0	0	-
87187/7	13	T	NA	0	40	20	40	0	-
87185/1	13	T	9.58	0	50	20	30	0	-
87185/2	13	T	4.5	0	10	0	90	0	-
87185/3	13	T	4.3	0	40	0	50	10	-
87185/4	13	T	6.7	0	20	10	50	20	-
88005/1	13	T	15.25	0	20	10	70	0	Amplified
88005/2	13	T	2.5	0	70	10	20	0	-

Table S 8

Description of the nodules isolated after treatment with SP600125 JNK inhibitor. Tumors are divided according to groups (T = treated; U = untreated). For each lesion shown are the age of the animal at the time of sacrifice, the size in mm of each lesion (for some lesions measurement could not be performed, N.A. = not applicable), and the histologic content. TaqMan copy number assay was done only on nodules between 10 mm and 20 mm in size to check whether the proportion of *Map2k7* amplications was comparable in treated and untreated animals, results are reported.

Lesion ID	Age (Years)	Gender	Conservation	Provenience	Background liver	% carcinoma
175	1.6	M	Frozen	LU	Extensive fibrosis	90%
7860	2.6	F	FFPE	BI	Mild fibrosis	90%
23836	1.3	M	FFPE	BI	Cirrhosis	90%
HB4R	8.6	F	Frozen	VF	Normal liver	70%
1790	11.7	M	Frozen	LU	Mild fibrosis	60%
2896	1.3	F	Frozen	LU	Cirrhosis	50%
MB	1.3	M	Frozen	TG	Cirrhosis	40%

Table S 9

PFIC samples description. For each lesion shown are the conservation of the sample, the provenience (BI = Hospital Papa Giovanni XXIII, Bergamo, Italy; LU = King's College Hospital, London, UK; VF = Hôpital Paul Brousse, Villejuif, France; TG = University of Tuebingen, Tuebingen, Germany), the histologic evaluation of the matched non-tumoral liver, and the percentage of carcinoma content.

	Tumor							Mean coverage of
Lesion ID	Samples/lane	Setting	Sequenced Gbps	Aligned Gbps	Aligned w/o duplicates	OnTarget Gbps	Mean Coverage	matched normal
175	2	101 PE	17.15	15.38	11.62	8.17	160	157
7860	1	76 PE	30.3	25.42	3.93	2.57	50	33
23836	2	76 PE	13.46	11.96	8.27	5.52	108	110
HB4R	2	101 PE	16.82	12.27	5.76	2.67	52	150
1790	2	101 PE	8.7	5.99	5.45	4.73	92	146
2896	2	101 PE	9.7	6.46	5.93	5.16	101	239

Table S 10

Sequencing settings, throughput and coverage for genomic analyses in PFIC HCCs. Details on the Illumina sequencing are provided, such as the number of samples per lane, the setting (PE = paired end sequencing), the sequenced gigabases (Gbps), the Gbps aligned to the genome, the Gbps after duplicate removal, the Gbps that were on targeted genomic regions, and the mean coverage of aligned reads used for variant calling in each sample. For the matched normal of each lesion, we followed the same experimental procedure of the tumor samples. Reported is only the mean coverage.

REFERENCES

1. Boveri T: **Concerning the origin of malignant tumours by Theodor Boveri. Translated and annotated by Henry Harris.** *Journal of cell science* 2008, **121 Suppl 1:**1-84.

2. Stehelin D, Varmus HE, Bishop JM, Vogt PK: **DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA.** *Nature* 1976, **260:**170-173.

3. Tabin CJ, Bradley SM, Bargmann CI, Weinberg RA, Papageorge AG, Scolnick EM, Dhar R, Lowy DR, Chang EH: **Mechanism of activation of a human oncogene.** *Nature* 1982, **300:**143-149.

4. Murphree AL, Benedict WF: **Retinoblastoma: clues to human oncogenesis.** *Science* 1984, **223:**1028-1033.

5. Guerrero I, Villasante A, Corces V, Pellicer A: **Activation of a c-K-ras oncogene by somatic mutation in mouse lymphomas induced by gamma radiation.** *Science* 1984, **225:**1159-1162.

6. Chorazy M: **Sequence rearrangements and genome instability. A possible step in carcinogenesis.** *Journal of cancer research and clinical oncology* 1985, **109:**159-172.

7. Hudson TJ, Anderson W, Artez A, Barker AD, Bell C, Bernabe RR, Bhan MK, Calvo F, Eerola I, Gerhard DS, et al: **International network of cancer genome projects.** *Nature* 2010, **464:**993-998.

8. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, et al: **Accurate whole human genome sequencing using reversible terminator chemistry.** *Nature* 2008, **456:**53-59.

9. Ley TJ, Mardis ER, Ding L, Fulton B, McLellan MD, Chen K, Dooling D, Dunford-Shore BH, McGrath S, Hickenbotham M, et al: **DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome.** *Nature* 2008, **456:**66-72.

10. Hodges E, Xuan Z, Balija V, Kramer M, Molla MN, Smith SW, Middle CM, Rodesch MJ, Albert TJ, Hannon GJ, McCombie WR: **Genome-wide in situ exon capture for selective resequencing.** *Nat Genet* 2007, **39:**1522-1527.

11. Network CGAR: **Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia.** *The New England journal of medicine* 2013, **368:**2059-2074.

12. Berger MF, Hodis E, Heffernan TP, Deribe YL, Lawrence MS, Protopopov A, Ivanova E, Watson IR, Nickerson E, Ghosh P, et al: **Melanoma genome sequencing reveals frequent PREX2 mutations.** *Nature* 2012, **485:**502-506.

13. Pleasance ED, Stephens PJ, O'Meara S, McBride DJ, Meynert A, Jones D, Lin ML, Beare D, Lau KW, Greenman C, et al: **A small-cell lung cancer genome with complex signatures of tobacco exposure.** *Nature* 2010, **463:**184-190.

14. Fujimoto A, Totoki Y, Abe T, Boroevich KA, Hosoda F, Nguyen HH, Aoki M, Hosono N, Kubo M, Miya F, et al: **Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators.** *Nature genetics* 2012, **44:**760-764.

15. Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, Jones D, Hinton J, Marshall J, Stebbings LA, et al: **Mutational processes molding the genomes of 21 breast cancers.** *Cell* 2012, **149:**979-993.

16. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Borresen-Dale AL, et al: **Signatures of mutational processes in human cancer.** *Nature* 2013, **500:**415-421.

17. Stephens PJ, Greenman CD, Fu B, Yang F, Bignell GR, Mudie LJ, Pleasance ED, Lau KW, Beare D, Stebbings LA, et al: **Massive genomic rearrangement acquired in a single catastrophic event during cancer development.** *Cell* 2011, **144:**27-40.

18. Crasta K, Ganem NJ, Dagher R, Lantermann AB, Ivanova EV, Pan Y, Nezi L, Protopopov A, Chowdhury D, Pellman D: **DNA breaks and chromosome pulverization from errors in mitosis.** *Nature* 2012, **482:**53-58.

19. Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, Barretina J, Boehm JS, Dobson J, Urashima M, et al: **The landscape of somatic copy-number alteration across human cancers.** *Nature* 2010, **463:**899-905.

20. Ciriello G, Miller ML, Aksoy BA, Senbabaoglu Y, Schultz N, Sander C: **Emerging landscape of oncogenic signatures across human cancers.** *Nature genetics* 2013, **45:**1127-1133.

21. Waddington CH: **Organisers and Genes.** *Cambridge Univ Press, Cambridge* 1940.

22. Murrell A, Rakyan VK, Beck S: **From genome to epigenome.** *Human molecular genetics* 2005, **14 Spec No 1:**R3-R10.

23. Feinberg AP, Ohlsson R, Henikoff S: **The epigenetic progenitor origin of human cancer.** *Nature reviews Genetics* 2006, **7:**21-33.

24. Feinberg AP, Vogelstein B: **Hypomethylation distinguishes genes of some human cancers from their normal counterparts.** *Nature* 1983, **301:**89-92.

25. Goelz SE, Vogelstein B, Hamilton SR, Feinberg AP: **Hypomethylation of DNA from benign and malignant human colon neoplasms.** *Science* 1985, **228:**187-190.

26. Robertson KD: **DNA methylation and human disease.** *Nature reviews Genetics* 2005, **6:**597-610.

27. Baylin SB, Ohm JE: **Epigenetic gene silencing in cancer - a mechanism for early oncogenic pathway addiction?** *Nature reviews Cancer* 2006, **6:**107-116.

28. Saxonov S, Berg P, Brutlag DL: **A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters.** *Proceedings of the National Academy of Sciences of the United States of America* 2006, **103:**1412-1417.

29. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, Cui H, Gabo K, Rongione M, Webster M, et al: **The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores.** *Nat Genet* 2009, **41:**178-186.

30. Doi A, Park IH, Wen B, Murakami P, Aryee MJ, Irizarry R, Herb B, Ladd-Acosta C, Rho J, Loewer S, et al: **Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts.** *Nature genetics* 2009, **41:**1350-1353.

31. Teschendorff AE, Jones A, Fiegl H, Sargent A, Zhuang JJ, Kitchener HC, Widschwendter M: **Epigenetic variability in cells of normal cytology is associated with the risk of future morphological transformation.** *Genome medicine* 2012, **4:**24.

32. Ebbs ML, Bartee L, Bender J: **H3 lysine 9 methylation is maintained on a transcribed inverted repeat by combined action of SUVH6 and SUVH4 methyltransferases.** *Molecular and cellular biology* 2005, **25:**10507-10515.

33. Jelinic P, Shaw P: **Loss of imprinting and cancer.** *The Journal of pathology* 2007, **211:**261-268.

34. Fournier C, Goto Y, Ballestar E, Delaval K, Hever AM, Esteller M, Feil R: **Allele-specific histone lysine methylation marks regulatory regions at imprinted mouse genes.** *The EMBO journal* 2002, **21:**6560-6570.

35. Ooi SK, Qiu C, Bernstein E, Li K, Jia D, Yang Z, Erdjument-Bromage H, Tempst P, Lin SP, Allis CD, et al: **DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA.** *Nature* 2007, **448:**714-717.

36. Henckel A, Nakabayashi K, Sanz LA, Feil R, Hata K, Arnaud P: **Histone methylation is mechanistically linked to DNA methylation at imprinting control regions in mammals.** *Human molecular genetics* 2009, **18:**3375-3383.

37. Cedar H, Bergman Y: **Linking DNA methylation and histone modification: patterns and paradigms.** *Nat Rev Genet* 2009, **10:**295-304.

38. Bachman KE, Park BH, Rhee I, Rajagopalan H, Herman JG, Baylin SB, Kinzler KW, Vogelstein B: **Histone modifications and silencing prior to DNA methylation of a tumor suppressor gene.** *Cancer Cell* 2003, **3:**89-95.

39. Fuks F: **DNA methylation and histone modifications: teaming up to silence genes.** *Current opinion in genetics & development* 2005, **15:**490-495.

40. Yu W, Gius D, Onyango P, Muldoon-Jacobs K, Karp J, Feinberg AP, Cui H: **Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA.** *Nature* 2008, **451:**202-206.

41. Timp W, Feinberg AP: **Cancer as a dysregulated epigenome allowing cellular growth advantage at the expense of the host.** *Nat Rev Cancer* 2013, **13:**497-510.

42. Sproul D, Nestor C, Culley J, Dickson JH, Dixon JM, Harrison DJ, Meehan RR, Sims AH, Ramsahoye BH: **Transcriptionally repressed genes become aberrantly methylated and distinguish tumors of different lineages in breast cancer.** *Proceedings of the National Academy of Sciences of the United States of America* 2011, **108:**4364-4369.

43. Sproul D, Kitchen RR, Nestor CE, Dixon JM, Sims AH, Harrison DJ, Ramsahoye BH, Meehan RR: **Tissue of origin determines cancer-associated CpG island promoter hypermethylation patterns.** *Genome biology* 2012, **13:**R84.

44. Fullgrabe J, Kavanagh E, Joseph B: **Histone onco-modifications.** *Oncogene* 2011, **30:**3391-3403.

45. Fraga MF, Ballestar E, Villar-Garea A, Boix-Chornet M, Espada J, Schotta G, Bonaldi T, Haydon C, Ropero S, Petrie K, et al: **Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer.** *Nature genetics* 2005, **37:**391-400.

46. Shogren-Knaak M, Ishii H, Sun JM, Pazin MJ, Davie JR, Peterson CL: **Histone H4-K16 acetylation controls chromatin structure and protein interactions.** *Science* 2006, **311:**844-847.

47. Schotta G, Lachner M, Sarma K, Ebert A, Sengupta R, Reuter G, Reinberg D, Jenuwein T: **A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin.** *Genes & development* 2004, **18:**1251-1262.

48. Seligson DB, Horvath S, Shi T, Yu H, Tze S, Grunstein M, Kurdistani SK: **Global histone modification patterns predict risk of prostate cancer recurrence.** *Nature* 2005, **435:**1262-1266.

49. Barlesi F, Giaccone G, Gallegos-Ruiz MI, Loundou A, Span SW, Lefesvre P, Kruyt FA, Rodriguez JA: **Global histone modifications predict prognosis of resected non small-cell lung cancer.** *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2007, **25:**4358-4364.

50. Elsheikh SE, Green AR, Rakha EA, Powe DG, Ahmed RA, Collins HM, Soria D, Garibaldi JM, Paish CE, Ammar AA, et al: **Global histone modifications in breast cancer correlate with tumor phenotypes, prognostic factors, and patient outcome.** *Cancer research* 2009, **69:**3802-3809.

51. Rajendran G, Shanmuganandam K, Bendre A, Muzumdar D, Goel A, Shiras A: **Epigenetic regulation of DNA methyltransferases: DNMT1 and DNMT3B in gliomas.** *Journal of neuro-oncology* 2011, **104:**483-494.

52. Krivtsov AV, Armstrong SA: **MLL translocations, histone modifications and leukaemia stem-cell development.** *Nature reviews Cancer* 2007, **7:**823-833.

53. Hamamoto R, Furukawa Y, Morita M, Iimura Y, Silva FP, Li M, Yagyu R, Nakamura Y: **SMYD3 encodes a histone methyltransferase involved in the proliferation of cancer cells.** *Nature cell biology* 2004, **6:**731-740.

54. Schulte JH, Lim S, Schramm A, Friedrichs N, Koster J, Versteeg R, Ora I, Pajtler K, Klein-Hitpass L, Kuhfittig-Kulle S, et al: **Lysine-specific demethylase 1 is strongly expressed in poorly differentiated neuroblastoma: implications for therapy.** *Cancer research* 2009, **69:**2065-2071.

55. Wang GG, Song J, Wang Z, Dormann HL, Casadio F, Li H, Luo JL, Patel DJ, Allis CD: **Haematopoietic malignancies caused by dysregulation of a chromatin-binding PHD finger.** *Nature* 2009, **459:**847-851.

56. Yu J, Rhodes DR, Tomlins SA, Cao X, Chen G, Mehra R, Wang X, Ghosh D, Shah RB, Varambally S, et al: **A polycomb repression signature in metastatic prostate cancer predicts cancer outcome.** *Cancer research* 2007, **67:**10657-10663.

57. Wei Y, Xia W, Zhang Z, Liu J, Wang H, Adsay NV, Albarracin C, Yu D, Abbruzzese JL, Mills GB, et al: **Loss of trimethylation at lysine 27 of histone H3 is a predictor of poor outcome in breast, ovarian, and pancreatic cancers.** *Molecular carcinogenesis* 2008, **47:**701-706.

58. van Haaften G, Dalgliesh GL, Davies H, Chen L, Bignell G, Greenman C, Edkins S, Hardy C, O'Meara S, Teague J, et al: **Somatic mutations of the histone H3K27 demethylase gene UTX in human cancer.** *Nature genetics* 2009, **41:**521-523.

59. Agger K, Cloos PA, Rudkjaer L, Williams K, Andersen G, Christensen J, Helin K: **The H3K27me3 demethylase JMJD3 contributes to the activation of the INK4A-ARF locus in response to oncogene- and stress-induced senescence.** *Genes & development* 2009, **23:**1171-1176.

60. Lewis PW, Muller MM, Koletsky MS, Cordero F, Lin S, Banaszynski LA, Garcia BA, Muir TW, Becher OJ, Allis CD: **Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma.** *Science* 2013, **340:**857-861.

61. Pogribny IP, Ross SA, Tryndyak VP, Pogribna M, Poirier LA, Karpinets TV: **Histone H3 lysine 9 and H4 lysine 20 trimethylation and the expression of Suv4-20h2 and Suv-39h1 histone methyltransferases in hepatocarcinogenesis induced by methyl deficiency in rats.** *Carcinogenesis* 2006, **27:**1180-1186.

62. Muller-Tidow C, Klein HU, Hascher A, Isken F, Tickenbrock L, Thoennissen N, Agrawal-Singh S, Tschanter P, Disselhoff C, Wang Y, et al: **Profiling of histone H3 lysine 9 trimethylation levels predicts transcription factor activity and survival in acute myeloid leukemia.** *Blood* 2010, **116:**3564-3571.

63. Kondo Y, Shen L, Suzuki S, Kurokawa T, Masuko K, Tanaka Y, Kato H, Mizuno Y, Yokoe M, Sugauchi F, et al: **Alterations of DNA methylation and histone modifications contribute to gene silencing in hepatocellular carcinomas.** *Hepatology research : the official journal of the Japan Society of Hepatology* 2007, **37:**974-983.

64. Cloos PA, Christensen J, Agger K, Maiolica A, Rappsilber J, Antal T, Hansen KH, Helin K: **The putative oncogene GASC1 demethylates tri- and dimethylated lysine 9 on histone H3.** *Nature* 2006, **442:**307-311.

65. Minucci S, Pelicci PG: **Histone deacetylase inhibitors and the promise of epigenetic (and more) treatments for cancer.** *Nature reviews Cancer* 2006, **6:**38-51.

66. Barker DJ: **The developmental origins of chronic adult disease.** *Acta paediatrica* 2004, **93:**26-33.

67. Gluckman PD, Hanson MA, Beedle AS: **Early life events and their consequences for later disease: a life history and evolutionary perspective.** *American journal of human biology : the official journal of the Human Biology Council* 2007, **19:**1-19.

68. Waterland RA, Jirtle RL: **Transposable elements: targets for early nutritional effects on epigenetic gene regulation.** *Molecular and cellular biology* 2003, **23:**5293-5300.

69. Ho SM, Tang WY, Belmonte de Frausto J, Prins GS: **Developmental exposure to estradiol and bisphenol A increases susceptibility to prostate carcinogenesis and epigenetically regulates phosphodiesterase type 4 variant 4.** *Cancer research* 2006, **66:**5624-5632.

70. Koturbash I, Baker M, Loree J, Kutanzi K, Hudson D, Pogribny I, Sedelnikova O, Bonner W, Kovalchuk O: **Epigenetic dysregulation underlies radiation-induced transgenerational genome instability in vivo.** *International journal of radiation oncology, biology, physics* 2006, **66:**327-330.

71. Jirtle RL, Skinner MK: **Environmental epigenomics and disease susceptibility.** *Nat Rev Genet* 2007, **8:**253-262.

72. Sachs L: **Hematopoietic growth and differentiation factors and the reversibility of malignancy: cell differentiation and by-passing of genetic defects in leukemia.** *Medical oncology and tumor pharmacotherapy* 1986, **3:**165-176.

73. Hochedlinger K, Blelloch R, Brennan C, Yamada Y, Kim M, Chin L, Jaenisch R: **Reprogramming of a melanoma genome by nuclear transplantation.** *Genes & development* 2004, **18:**1875-1885.

74. Bussard KM, Smith GH: **Human breast cancer cells are redirected to mammary epithelial cells upon interaction with the regenerating mammary gland microenvironment in-vivo.** *PLoS One* 2012, **7:**e49221.

75. Kulesa PM, Kasemeier-Kulesa JC, Teddy JM, Margaryan NV, Seftor EA, Seftor RE, Hendrix MJ: **Reprogramming metastatic melanoma cells to assume a neural crest cell-like phenotype in an embryonic microenvironment.** *Proceedings of the National Academy of Sciences of the United States of America* 2006, **103:**3752-3757.

76. Sakatani T, Kaneda A, Iacobuzio-Donahue CA, Carter MG, de Boom Witzel S, Okano H, Ko MS, Ohlsson R, Longo DL, Feinberg AP: **Loss of imprinting of Igf2 alters intestinal maturation and tumorigenesis in mice.** *Science* 2005, **307:**1976-1978.

77. Alvarez H, Opalinska J, Zhou L, Sohal D, Fazzari MJ, Yu Y, Montagna C, Montgomery EA, Canto M, Dunbar KB, et al: **Widespread hypomethylation occurs early and synergizes with gene amplification during esophageal carcinogenesis.** *PLoS genetics* 2011, **7:**e1001356.

78. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S, Klotzle B, Bibikova M, Fan JB, Gao Y, et al: **Genome-wide methylation profiles reveal quantitative views of human aging rates.** *Molecular cell* 2013, **49:**359-367.

79. Versteege I, Sevenet N, Lange J, Rousseau-Merck MF, Ambros P, Handgretinger R, Aurias A, Delattre O: **Truncating mutations of hSNF5/INI1 in aggressive paediatric cancer.** *Nature* 1998, **394:**203-206.

80. Pugh TJ, Weeraratne SD, Archer TC, Pomeranz Krummel DA, Auclair D, Bochicchio J, Carneiro MO, Carter SL, Cibulskis K, Erlich RL, et al: **Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations.** *Nature* 2012, **488:**106-110.

81. Gui Y, Guo G, Huang Y, Hu X, Tang A, Gao S, Wu R, Chen C, Li X, Zhou L, et al: **Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder.** *Nature genetics* 2011, **43:**875-878.

82. Wiegand KC, Shah SP, Al-Agha OM, Zhao Y, Tse K, Zeng T, Senz J, McConechy MK, Anglesio MS, Kalloger SE, et al: **ARID1A mutations in endometriosis-associated ovarian carcinomas.** *The New England journal of medicine* 2010, **363:**1532-1543.

83. Schwartzentruber J, Korshunov A, Liu XY, Jones DT, Pfaff E, Jacob K, Sturm D, Fontebasso AM, Quang DA, Tonjes M, et al: **Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma.** *Nature* 2012, **482:**226-231.

84. Jones S, Zhang X, Parsons DW, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H, Kamiyama H, Jimeno A, et al: **Core signaling pathways in human pancreatic cancers revealed by global genomic analyses.** *Science* 2008, **321:**1801-1806.

85. Yan XJ, Xu J, Gu ZH, Pan CM, Lu G, Shen Y, Shi JY, Zhu YM, Tang L, Zhang XW, et al: **Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia.** *Nature genetics* 2011, **43:**309-315.

86. Figueroa ME, Abdel-Wahab O, Lu C, Ward PS, Patel J, Shih A, Li Y, Bhagwat N, Vasanthakumar A, Fernandez HF, et al: **Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation.** *Cancer Cell* 2010, **18:**553-567.

87. Pasqualucci L, Dominguez-Sola D, Chiarenza A, Fabbri G, Grunn A, Trifonov V, Kasper LH, Lerach S, Tang H, Ma J, et al: **Inactivating mutations of acetyltransferase genes in B-cell lymphoma.** *Nature* 2011, **471:**189-195.

88. Li J, Harris RA, Cheung SW, Coarfa C, Jeong M, Goodell MA, White LD, Patel A, Kang SH, Shaw C, et al: **Genomic hypomethylation in the human germline associates with selective structural mutability in the human genome.** *PLoS genetics* 2012, **8:**e1002692.

89. Cadieux B, Ching TT, VandenBerg SR, Costello JF: **Genome-wide hypomethylation in human glioblastomas associated with specific copy number alteration, methylenetetrahydrofolate reductase allele status, and increased proliferation.** *Cancer research* 2006, **66:**8469-8476.

90. De S, Michor F: **DNA replication timing and long-range DNA interactions predict mutational landscapes of cancer genomes.** *Nat Biotechnol* 2011, **29:**1103-1108.

91. Hodgkinson A, Eyre-Walker A: **Variation in the mutation rate across mammalian genomes.** *Nature reviews Genetics* 2011, **12:**756-766.

92. Olivier M, Hollstein M, Hainaut P: **TP53 mutations in human cancers: origins, consequences, and clinical use.** *Cold Spring Harbor perspectives in biology* 2010, **2:**a001008.

93. Schuster-Bockler B, Lehner B: **Chromatin organization is a major influence on regional mutation rates in human cancer cells.** *Nature* 2012, **488:**504-507.

94. Jiang Y, Lucas I, Young DJ, Davis EM, Karrison T, Rest JS, Le Beau MM: **Common fragile sites are characterized by histone hypoacetylation.** *Human molecular genetics* 2009, **18:**4501-4512.

95. Toyota M, Suzuki H: **Epigenetic drivers of genetic alterations.** *Advances in genetics* 2010, **70:**309-323.

96. De S, Michor F: **DNA secondary structures and epigenetic determinants of cancer genome evolution.** *Nature structural & molecular biology* 2011, **18:**950-955.

97. Jones PA, Baylin SB: **The fundamental role of epigenetic events in cancer.** *Nature reviews Genetics* 2002, **3:**415-428.

98. Graff JR, Herman JG, Lapidus RG, Chopra H, Xu R, Jarrard DF, Isaacs WB, Pitha PM, Davidson NE, Baylin SB: **E-cadherin expression is silenced by DNA hypermethylation in human breast and prostate carcinomas.** *Cancer research* 1995, **55:**5195-5199.

99. Spellman PT: **Integrated genomic analyses of ovarian carcinoma.** *Nature* 2011, **474:**609-615.

100. Meyerson M: **Comprehensive genomic characterization of squamous cell lung cancers.** *Nature* 2012, **489:**519-525.

101. Hinoue T, Weisenberger DJ, Pan F, Campan M, Kim M, Young J, Whitehall VL, Leggett BA, Laird PW: **Analysis of the association between CIMP and BRAF in colorectal cancer by DNA methylation profiling.** *PLoS One* 2009, **4:**e8357.

102. Dvorak HF: **Tumors: wounds that do not heal. Similarities between tumor stroma generation and wound healing.** *The New England journal of medicine* 1986, **315:**1650-1659.

103. Osborn O, Olefsky JM: **The cellular and signaling networks linking the immune system and metabolism in disease.** *Nature medicine* 2012, **18:**363-374.

104. Mantovani A: **Inflammation and cancer: the macrophage connection.** *Medicina (Buenos Aires)* 2007, **67:**32-34.

105. Aggarwal BB, Vijayalekshmi RV, Sung B: **Targeting inflammatory pathways for prevention and therapy of cancer: short-term friend, long-term foe.** *Clinical cancer research : an official journal of the American Association for Cancer Research* 2009, **15:**425-430.

106. Mantovani A, Allavena P, Sica A, Balkwill F: **Cancer-related inflammation.** *Nature* 2008, **454:**436-444.

107. Wang D, DuBois RN: **The role of anti-inflammatory drugs in colorectal cancer.** *Annual review of medicine* 2013, **64:**131-144.

108. Grivennikov SI, Greten FR, Karin M: **Immunity, inflammation, and cancer.** *Cell* 2010, **140:**883-899.

109. Lin WW, Karin M: **A cytokine-mediated link between innate immunity, inflammation, and cancer.** *J Clin Invest* 2007, **117:**1175-1183.

110. DeNardo DG, Andreu P, Coussens LM: **Interactions between lymphocytes and myeloid cells regulate pro- versus anti-tumor immunity.** *Cancer metastasis reviews* 2010, **29:**309-316.

111. Qian BZ, Pollard JW: **Macrophage diversity enhances tumor progression and metastasis.** *Cell* 2010, **141:**39-51.

112. Hanahan D, Weinberg RA: **Hallmarks of cancer: the next generation.** *Cell* 2011, **144:**646-674.

113. Sato Y, Takahashi S, Kinouchi Y, Shiraki M, Endo K, Matsumura Y, Kakuta Y, Tosa M, Motida A, Abe H, et al: **IL-10 deficiency leads to somatic mutations in a model of IBD.** *Carcinogenesis* 2006, **27:**1068-1073.

114. Bielas JH, Loeb KR, Rubin BP, True LD, Loeb LA: **Human cancers express a mutator phenotype.** *Proceedings of the National Academy of Sciences of the United States of America* 2006, **103:**18238-18242.

115. Hussain SP, Amstad P, Raja K, Ambs S, Nagashima M, Bennett WP, Shields PG, Ham AJ, Swenberg JA, Marrogi AJ, Harris CC: **Increased p53 mutation load in noncancerous colon tissue from ulcerative colitis: a cancer-prone chronic inflammatory disease.** *Cancer research* 2000, **60:**3333-3337.

116. Firestein GS, Echeverri F, Yeo M, Zvaifler NJ, Green DR: **Somatic mutations in the p53 tumor suppressor gene in rheumatoid arthritis synovium.** *Proceedings of the National Academy of Sciences of the United States of America* 1997, **94:**10895-10900.

117. Andreassi MG, Botto N: **Genetic instability, DNA damage and atherosclerosis.** *Cell Cycle* 2003, **2:**224-227.

118. Danish M, Sissung T, Price DK, Figg WD: **Genetic stability of tumor microenvironment.** *Cancer biology & therapy* 2008, **7:**331-332.

119. Polyak K, Haviv I, Campbell IG: **Co-evolution of tumor cells and their microenvironment.** *Trends in genetics : TIG* 2009, **25:**30-38.

120. Qiu W, Hu M, Sridhar A, Opeskin K, Fox S, Shipitsin M, Trivett M, Thompson ER, Ramakrishna M, Gorringe KL, et al: **No evidence of clonal somatic genetic alterations in cancer-associated fibroblasts from human breast and ovarian carcinomas.** *Nature genetics* 2008, **40:**650-655.

121. Colotta F, Allavena P, Sica A, Garlanda C, Mantovani A: **Cancer-related inflammation, the seventh hallmark of cancer: links to genetic instability.** *Carcinogenesis* 2009, **30:**1073-1081.

122. Matsumoto Y, Marusawa H, Kinoshita K, Endo Y, Kou T, Morisawa T, Azuma T, Okazaki IM, Honjo T, Chiba T: **Helicobacter pylori infection triggers aberrant expression of activation-induced cytidine deaminase in gastric epithelium.** *Nature medicine* 2007, **13:**470-476.

123. Endo Y, Marusawa H, Kou T, Nakase H, Fujii S, Fujimori T, Kinoshita K, Honjo T, Chiba T: **Activation-induced cytidine deaminase links between inflammation and the development of colitis-associated colorectal cancers.** *Gastroenterology* 2008, **135:**889-898, 898 e881-883.

124. Kou T, Marusawa H, Kinoshita K, Endo Y, Okazaki IM, Ueda Y, Kodama Y, Haga H, Ikai I, Chiba T: **Expression of activation-induced cytidine deaminase in human hepatocytes during hepatocarcinogenesis.** *International journal of cancer Journal international du cancer* 2007, **120:**469-476.

125. Robbiani DF, Bothmer A, Callen E, Reina-San-Martin B, Dorsett Y, Difilippantonio S, Bolland DJ, Chen HT, Corcoran AE, Nussenzweig A, Nussenzweig MC: **AID is required for the chromosomal breaks in c-myc that lead to c-myc/IgH translocations.** *Cell* 2008, **135:**1028-1038.

126. Shen HM, Peters A, Baron B, Zhu X, Storb U: **Mutation of BCL-6 gene in normal B cells by the process of somatic hypermutation of Ig genes.** *Science* 1998, **280:**1750-1752.

127. Myant KB, Cammareri P, McGhee EJ, Ridgway RA, Huels DJ, Cordero JB, Schwitalla S, Kalna G, Ogg EL, Athineos D, et al: **ROS production and NF-kappaB activation triggered by RAC1 facilitate WNT-driven intestinal stem cell proliferation and colorectal cancer initiation.** *Cell Stem Cell* 2013, **12:**761-773.

128. Ito K, Barnes PJ, Adcock IM: **Glucocorticoid receptor recruitment of histone deacetylase 2 inhibits interleukin-1beta-induced histone H4 acetylation on lysines 8 and 12.** *Molecular and cellular biology* 2000, **20:**6891-6903.

129. Valinluck V, Sowers LC: **Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1.** *Cancer Res* 2007, **67:**946-950.

130. Kondo Y, Kanai Y, Sakamoto M, Mizokami M, Ueda R, Hirohashi S: **Genetic instability and aberrant DNA methylation in chronic hepatitis and cirrhosis--A comprehensive study of loss of heterozygosity and microsatellite instability at 39 loci and DNA hypermethylation on 8 CpG islands in microdissected specimens from patients with hepatocellular carcinoma.** *Hepatology* 2000, **32:**970-979.

131. Stenvinkel P, Karimi M, Johansson S, Axelsson J, Suliman M, Lindholm B, Heimburger O, Barany P, Alvestrand A, Nordfors L, et al: **Impact of inflammation on epigenetic DNA methylation - a novel risk factor for cardiovascular disease?** *J Intern Med* 2007, **261:**488-499.

132. Lund G, Andersson L, Lauria M, Lindholm M, Fraga MF, Villar-Garea A, Ballestar E, Esteller M, Zaina S: **DNA methylation polymorphisms precede any histological sign of atherosclerosis in mice lacking apolipoprotein E.** *The Journal of biological chemistry* 2004, **279:**29147-29154.

133. Hodge DR, Xiao W, Clausen PA, Heidecker G, Szyf M, Farrar WL: **Interleukin-6 regulation of the human DNA methyltransferase (HDNMT) gene in human erythroleukemia cells.** *J Biol Chem* 2001, **276:**39508-39511.

134. Wehbe H, Henson R, Meng F, Mize-Berge J, Patel T: **Interleukin-6 contributes to growth in cholangiocarcinoma cells by aberrant promoter methylation and gene expression.** *Cancer Res* 2006, **66:**10517-10524.

135. Gasche JA, Hoffmann J, Boland CR, Goel A: **Interleukin-6 promotes tumorigenesis by altering DNA methylation in oral cancer cells.** *International journal of cancer Journal international du cancer* 2011, **129:**1053-1063.

136. Hahn MA, Hahn T, Lee DH, Esworthy RS, Kim BW, Riggs AD, Chu FF, Pfeifer GP: **Methylation of polycomb target genes in intestinal cancer is mediated by inflammation.** *Cancer Res* 2008, **68:**10280-10289.

137. De Santa F, Totaro MG, Prosperini E, Notarbartolo S, Testa G, Natoli G: **The histone H3 lysine-27 demethylase Jmjd3 links inflammation to inhibition of polycomb-mediated gene silencing.** *Cell* 2007, **130:**1083-1094.

138. Takeshima H, Ikegami D, Wakabayashi M, Niwa T, Kim YJ, Ushijima T: **Induction of aberrant trimethylation of histone H3 lysine 27 by inflammation in mouse colonic epithelial cells.** *Carcinogenesis* 2012, **33:**2384-2390.

139. **GLOBOCAN 2008 v2.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 10 [Internet].** [http://globocan.iarc.fr, accessed on day/month/year.]

140. Spangenberg HC, Thimme R, Blum HE: **Targeted therapy for hepatocellular carcinoma.** *Nature reviews Gastroenterology & hepatology* 2009, **6:**423-432.

141. Hernandez-Gea V, Toffanin S, Friedman SL, Llovet JM: **Role of the microenvironment in the pathogenesis and treatment of hepatocellular carcinoma.** *Gastroenterology* 2013, **144:**512-527.

142. Page JM, Harrison SA: **NASH and HCC.** *Clinics in liver disease* 2009, **13:**631-647.

143. Baffy G, Brunt EM, Caldwell SH: **Hepatocellular carcinoma in non-alcoholic fatty liver disease: an emerging menace.** *Journal of hepatology* 2012, **56:**1384-1391.

144. Lohse AW, Mieli-Vergani G: **Autoimmune hepatitis.** *Journal of hepatology* 2011, **55:**171-182.

145. El-Serag HB, Tran T, Everhart JE: **Diabetes increases the risk of chronic liver disease and hepatocellular carcinoma.** *Gastroenterology* 2004, **126:**460-468.

146. Dragani TA: **Risk of HCC: genetic heterogeneity and complex genetics.** *Journal of hepatology* 2010, **52:**252-257.

147. Sun B, Karin M: **Obesity, inflammation, and liver cancer.** *Journal of hepatology* 2012, **56:**704-713.

148. Ribes J, Cleries R, Esteban L, Moreno V, Bosch FX: **The influence of alcohol consumption and hepatitis B and C infections on the risk of liver cancer in Europe.** *Journal of hepatology* 2008, **49:**233-242.

149. Tao Y, Ruan J, Yeh SH, Lu X, Wang Y, Zhai W, Cai J, Ling S, Gong Q, Chong Z, et al: **Rapid growth of a hepatocellular carcinoma and the driving mutations revealed by cell-population genetic analysis of whole-genome data.** *Proceedings of the National Academy of Sciences of the United States of America* 2011, **108:**12042-12047.

150. Totoki Y, Tatsuno K, Yamamoto S, Arai Y, Hosoda F, Ishikawa S, Tsutsumi S, Sonoda K, Totsuka H, Shirakihara T, et al: **High-resolution characterization of a hepatocellular carcinoma genome.** *Nature genetics* 2011, **43:**464-469.

151. Li M, Zhao H, Zhang X, Wood LD, Anders RA, Choti MA, Pawlik TM, Daniel HD, Kannangai R, Offerhaus GJ, et al: **Inactivating mutations of the chromatin remodeling gene ARID2 in hepatocellular carcinoma.** *Nat Genet* 2011, **43:**828-829.

152. Guichard C, Amaddeo G, Imbeaud S, Ladeiro Y, Pelletier L, Maad IB, Calderaro J, Bioulac-Sage P, Letexier M, Degos F, et al: **Integrated analysis of somatic mutations and**

focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. *Nat Genet* 2012, **44:**694-698.

153. Sung WK, Zheng H, Li S, Chen R, Liu X, Li Y, Lee NP, Lee WH, Ariyaratne PN, Tennakoon C, et al: **Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma.** *Nat Genet* 2012, **44:**765-769.

154. Huang J, Deng Q, Wang Q, Li KY, Dai JH, Li N, Zhu ZD, Zhou B, Liu XY, Liu RF, et al: **Exome sequencing of hepatitis B virus-associated hepatocellular carcinoma.** *Nat Genet* 2012, **44:**1117-1121.

155. Kan Z, Zheng H, Liu X, Li S, Barber TD, Gong Z, Gao H, Hao K, Willard MD, Xu J, et al: **Whole-genome sequencing identifies recurrent mutations in hepatocellular carcinoma.** *Genome Res* 2013.

156. Hsu IC, Metcalf RA, Sun T, Welsh JA, Wang NJ, Harris CC: **Mutational hotspot in the p53 gene in human hepatocellular carcinomas.** *Nature* 1991, **350:**427-428.

157. Smit JJ, Schinkel AH, Oude Elferink RP, Groen AK, Wagenaar E, van Deemter L, Mol CA, Ottenhoff R, van der Lugt NM, van Roon MA, et al.: **Homozygous disruption of the murine mdr2 P-glycoprotein gene leads to a complete absence of phospholipid from bile and to liver disease.** *Cell* 1993, **75:**451-462.

158. Mauad TH, van Nieuwkerk CM, Dingemans KP, Smit JJ, Schinkel AH, Notenboom RG, van den Bergh Weerman MA, Verkruisen RP, Groen AK, Oude Elferink RP, et al.: **Mice with homozygous disruption of the mdr2 P-glycoprotein gene. A novel animal model for studies of nonsuppurative inflammatory cholangitis and hepatocarcinogenesis.** *Am J Pathol* 1994, **145:**1237-1245.

159. Jacquemin E: **Role of multidrug resistance 3 deficiency in pediatric and adult liver disease: one gene for three diseases.** *Seminars in liver disease* 2001, **21:**551-562.

160. Pikarsky E, Porat RM, Stein I, Abramovitch R, Amit S, Kasem S, Gutkovich-Pyest E, Urieli-Shoval S, Galun E, Ben-Neriah Y: **NF-kappaB functions as a tumour promoter in inflammation-associated cancer.** *Nature* 2004, **431:**461-466.

161. Barashi N, Weiss ID, Wald O, Wald H, Beider K, Abraham M, Klein S, Goldenberg D, Axelrod J, Pikarsky E, et al: **Inflammation-induced hepatocellular carcinoma is dependent on CCR5 in mice.** *Hepatology* 2013, **58:**1021-1030.

162. Katzenellenbogen M, Mizrahi L, Pappo O, Klopstock N, Olam D, Jacob-Hirsch J, Amariglio N, Rechavi G, Domany E, Galun E, Goldenberg D: **Molecular mechanisms of liver carcinogenesis in the mdr2-knockout mice.** *Mol Cancer Res* 2007, **5:**1159-1170.

163. Jang JJ, Weghorst CM, Henneman JR, Devor DE, Ward JM: **Progressive atypia in spontaneous and N-nitrosodiethylamine-induced hepatocellular adenomas of C3H/HeNCr mice.** *Carcinogenesis* 1992, **13:**1541-1547.

164. Oude Elferink RP, Paulusma CC: **Function and pathophysiological importance of ABCB4 (MDR3 P-glycoprotein).** *Pflugers Archiv : European journal of physiology* 2007, **453:**601-610.

165. Davit-Spraul A, Gonzales E, Baussan C, Jacquemin E: **Progressive familial intrahepatic cholestasis.** *Orphanet journal of rare diseases* 2009, **4:**1.

166. Nicolaou M, Andress EJ, Zolnerciks JK, Dixon PH, Williamson C, Linton KJ: **Canalicular ABC transporters and liver disease.** *The Journal of pathology* 2012, **226:**300-315.

167. Paulusma CC, Folmer DE, Ho-Mok KS, de Waart DR, Hilarius PM, Verhoeven AJ, Oude Elferink RP: **ATP8B1 requires an accessory protein for endoplasmic reticulum exit and plasma membrane lipid flippase activity.** *Hepatology* 2008, **47:**268-278.

168. Clayton RJ, Iber FL, Ruebner BH, McKusick VA: **Byler disease. Fatal familial intrahepatic cholestasis in an Amish kindred.** *American journal of diseases of children* 1969, **117:**112-124.

169. Chen F, Ananthanarayanan M, Emre S, Neimark E, Bull LN, Knisely AS, Strautnieks SS, Thompson RJ, Magid MS, Gordon R, et al: **Progressive familial intrahepatic cholestasis, type 1, is associated with decreased farnesoid X receptor activity.** *Gastroenterology* 2004, **126:**756-764.

170. Bull LN, Carlton VE, Stricker NL, Baharloo S, DeYoung JA, Freimer NB, Magid MS, Kahn E, Markowitz J, DiCarlo FJ, et al: **Genetic and morphological findings in progressive familial intrahepatic cholestasis (Byler disease [PFIC-1] and Byler syndrome): evidence for heterogeneity.** *Hepatology* 1997, **26:**155-164.

171. Bull LN, van Eijk MJ, Pawlikowska L, DeYoung JA, Juijn JA, Liao M, Klomp LW, Lomri N, Berger R, Scharschmidt BF, et al: **A gene encoding a P-type ATPase mutated in two forms of hereditary cholestasis.** *Nature genetics* 1998, **18:**219-224.

172. Stieger B: **Recent insights into the function and regulation of the bile salt export pump (ABCB11).** *Current opinion in lipidology* 2009, **20:**176-181.

173. Kubitz R, Droge C, Stindt J, Weissenberger K, Haussinger D: **The bile salt export pump (BSEP) in health and disease.** *Clinics and research in hepatology and gastroenterology* 2012, **36:**536-553.

174. Evason K, Bove KE, Finegold MJ, Knisely AS, Rhee S, Rosenthal P, Miethke AG, Karpen SJ, Ferrell LD, Kim GE: **Morphologic findings in progressive familial intrahepatic cholestasis 2 (PFIC2): correlation with genetic and immunohistochemical studies.** *The American journal of surgical pathology* 2011, **35:**687-696.

175. Sheridan RM, Gupta A, Miethke A, Knisely AS, Bove KE: **Multiple dysplastic liver nodules in PFIC2 underscore risk for neoplasia associated with functional BSEP deficiency.** *The American journal of surgical pathology* 2012, **36:**785-786.

176. Romano F, Stroppa P, Bravi M, Casotti V, Lucianetti A, Guizzetti M, Sonzogni A, Colledan M, D'Antiga L: **Favorable outcome of primary liver transplantation in children with cirrhosis and hepatocellular carcinoma.** *Pediatric transplantation* 2011, **15:**573-579.

177. Knisely AS, Strautnieks SS, Meier Y, Stieger B, Byrne JA, Portmann BC, Bull LN, Pawlikowska L, Bilezikci B, Ozcay F, et al: **Hepatocellular carcinoma in ten children under five years of age with bile salt export pump deficiency.** *Hepatology* 2006, **44:**478-486.

178. Gros P, Raymond M, Bell J, Housman D: **Cloning and characterization of a second member of the mouse mdr gene family.** *Molecular and cellular biology* 1988, **8:**2770-2778.

179. Smith AJ, de Vree JM, Ottenhoff R, Oude Elferink RP, Schinkel AH, Borst P: **Hepatocyte-specific expression of the human MDR3 P-glycoprotein gene restores the biliary phosphatidylcholine excretion absent in Mdr2 (-/-) mice.** *Hepatology* 1998, **28:**530-536.

180. Chan J, Vandeberg JL: **Hepatobiliary transport in health and disease.** *Clinical lipidology* 2012, **7:**189-202.

181. Wang R, Salem M, Yousef IM, Tuchweber B, Lam P, Childs SJ, Helgason CD, Ackerley C, Phillips MJ, Ling V: **Targeted inactivation of sister of P-glycoprotein gene (spgp) in mice results in nonprogressive but persistent intrahepatic cholestasis.** *Proceedings of the National Academy of Sciences of the United States of America* 2001, **98:**2011-2016.

182. Wang R, Lam P, Liu L, Forrest D, Yousef IM, Mignault D, Phillips MJ, Ling V: **Severe cholestasis induced by cholic acid feeding in knockout mice of sister of P-glycoprotein.** *Hepatology* 2003, **38:**1489-1499.

183. Lam P, Wang R, Ling V: **Bile acid transport in sister of P-glycoprotein (ABCB11) knockout mice.** *Biochemistry* 2005, **44:**12598-12605.

184. Keitel V, Burdelski M, Warskulat U, Kuhlkamp T, Keppler D, Haussinger D, Kubitz R: **Expression and localization of hepatobiliary transport proteins in progressive familial intrahepatic cholestasis.** *Hepatology* 2005, **41:**1160-1172.

185. Basso K, Margolin AA, Stolovitzky G, Klein U, Dalla-Favera R, Califano A: **Reverse engineering of regulatory networks in human B cells.** *Nature genetics* 2005, **37:**382-390.

186. Carro MS, Lim WK, Alvarez MJ, Bollo RJ, Zhao X, Snyder EY, Sulman EP, Anne SL, Doetsch F, Colman H, et al: **The transcriptional network for mesenchymal transformation of brain tumours.** *Nature* 2010, **463:**318-325.

187. Berry MN, Friend DS: **High-yield preparation of isolated rat liver parenchymal cells: a biochemical and fine structural study.** *The Journal of cell biology* 1969, **43:**506-520.

188. Seglen PO: **Preparation of isolated rat liver cells.** *Methods Cell Biol* 1976, **13:**29-83.

189. Thoolen B, Maronpot RR, Harada T, Nyska A, Rousseaux C, Nolte T, Malarkey DE, Kaufmann W, Kuttler K, Deschl U, et al: **Proliferative and nonproliferative lesions of the rat and mouse hepatobiliary system.** *Toxicologic pathology* 2010, **38:**5S-81S.

190. Ghisletti S, Barozzi I, Mietton F, Polletti S, De Santa F, Venturini E, Gregory L, Lonie L, Chew A, Wei CL, et al: **Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages.** *Immunity* 2010, **32:**317-328.

191. Barozzi I, Termanini A, Minucci S, Natoli G: **Fish the ChIPs: a pipeline for automated genomic annotation of ChIP-Seq data.** *Biology direct* 2011, **6:**51.

192. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS: **Model-based analysis of ChIP-Seq (MACS).** *Genome biology* 2008, **9:**R137.

193. Abdi H, Williams LJ: **Principal component analysis.** *WIREs Comp Stat* 2010, **2:**433-459.

194. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G: **GREAT improves functional interpretation of cis-regulatory regions.** *Nature biotechnology* 2010, **28:**495-501.

195. Zambelli F, Pesole G, Pavesi G: **Pscan: finding over-represented transcription factor binding site motifs in sequences from co-regulated or co-expressed genes.** *Nucleic acids research* 2009, **37:**W247-252.

196. Portales-Casamar E, Thongjuea S, Kwon AT, Arenillas D, Zhao X, Valen E, Yusuf D, Lenhard B, Wasserman WW, Sandelin A: **JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles.** *Nucleic acids research* 2010, **38:**D105-110.

197. Badis G, Berger MF, Philippakis AA, Talukder S, Gehrke AR, Jaeger SA, Chan ET, Metzler G, Vedenko A, Chen X, et al: **Diversity and complexity in DNA recognition by transcription factors.** *Science* 2009, **324:**1720-1723.

198. Berger MF, Badis G, Gehrke AR, Talukder S, Philippakis AA, Pena-Castillo L, Alleyne TM, Mnaimneh S, Botvinnik OB, Chan ET, et al: **Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences.** *Cell* 2008, **133:**1266-1276.

199. Bucher P: **Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences.** *Journal of molecular biology* 1990, **212:**563-578.

200. Wei GH, Badis G, Berger MF, Kivioja T, Palin K, Enge M, Bonke M, Jolma A, Varjosalo M, Gehrke AR, et al: **Genome-wide analysis of ETS-family DNA-binding in vitro and in vivo.** *The EMBO journal* 2010, **29:**2147-2160.

201. Grant CE, Bailey TL, Noble WS: **FIMO: scanning for occurrences of a given motif.** *Bioinformatics* 2011, **27:**1017-1018.

202. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R: **The Sequence Alignment/Map format and SAMtools.** *Bioinformatics* 2009, **25:**2078-2079.

203. Carter SL, Cibulskis K, Helman E, McKenna A, Shen H, Zack T, Laird PW, Onofrio RC, Winckler W, Weir BA, et al: **Absolute quantification of somatic DNA alterations in human cancer.** *Nat Biotechnol* 2012, **30:**413-421.

204. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK: **VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing.** *Genome research* 2012, **22:**568-576.

205. Abyzov A, Urban AE, Snyder M, Gerstein M: **CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing.** *Genome Res* 2011, **21:**974-984.

206. D'Antonio M, Pendino V, Sinha S, Ciccarelli FD: **Network of Cancer Genes (NCG 3.0): integration and analysis of genetic and network properties of cancer genes.** *Nucleic Acids Res* 2012, **40:**D978-983.

207. Eppig JT, Blake JA, Bult CJ, Kadin JA, Richardson JE: **The Mouse Genome Database (MGD): comprehensive resource for genetics and genomics of the laboratory mouse.** *Nucleic acids research* 2012, **40:**D881-886.

208. Muller J, Szklarczyk D, Julien P, Letunic I, Roth A, Kuhn M, Powell S, von Mering C, Doerks T, Jensen LJ, Bork P: **eggNOG v2.0: extending the evolutionary genealogy of genes with enhanced non-supervised orthologous groups, species and functional annotations.** *Nucleic acids research* 2010, **38:**D190-195.

209. Zhang Z: **Genomic landscape of liver cancer.** *Nature genetics* 2012, **44:**1075-1077.

210. Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, Rahman N, Stratton MR: **A census of human cancer genes.** *Nat Rev Cancer* 2004, **4:**177-183.

211. Ge X, Yamamoto S, Tsutsumi S, Midorikawa Y, Ihara S, Wang SM, Aburatani H: **Interpreting expression profiles of cancers by genome-wide survey of breadth of expression in normal tissues.** *Genomics* 2005, **86:**127-141.

212. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, Zhang J, Soden R, Hayakawa M, Kreiman G, et al: **A gene atlas of the mouse and human protein-encoding transcriptomes.** *Proceedings of the National Academy of Sciences of the United States of America* 2004, **101:**6062-6067.

213. Gautier L, Cope L, Bolstad BM, Irizarry RA: **affy--analysis of Affymetrix GeneChip data at the probe level.** *Bioinformatics* 2004, **20:**307-315.

214. Van Loo P, Nordgard SH, Lingjaerde OC, Russnes HG, Rye IH, Sun W, Weigman VJ, Marynen P, Zetterberg A, Naume B, et al: **Allele-specific copy number analysis of tumors.** *Proceedings of the National Academy of Sciences of the United States of America* 2010, **107:**16910-16915.

215. Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, et al: **Histone H3K27ac separates active from poised enhancers and predicts developmental state.** *Proceedings of the National Academy of Sciences of the United States of America* 2010, **107:**21931-21936.

216. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, et al: **Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome.** *Nature genetics* 2007, **39:**311-318.

217. Karin M, Liu Z, Zandi E: **AP-1 function and regulation.** *Current opinion in cell biology* 1997, **9:**240-246.

218. Hess J, Angel P, Schorpp-Kistner M: **AP-1 subunits: quarrel and harmony among siblings.** *Journal of cell science* 2004, **117:**5965-5973.

219. Seth A, Watson DK: **ETS transcription factors and their emerging roles in human cancer.** *European journal of cancer* 2005, **41:**2462-2478.

220. Gallant S, Gilkeson G: **ETS transcription factors and regulation of immunity.** *Archivum immunologiae et therapiae experimentalis* 2006, **54:**149-163.

221. Lilljebjorn H, Rissler M, Lassen C, Heldrup J, Behrendtz M, Mitelman F, Johansson B, Fioretos T: **Whole-exome sequencing of pediatric acute lymphoblastic leukemia.** *Leukemia* 2012, **26:**1602-1607.

222. Piazza R, Valletta S, Winkelmann N, Redaelli S, Spinelli R, Pirola A, Antolini L, Mologni L, Donadoni C, Papaemmanuil E, et al: **Recurrent SETBP1 mutations in atypical chronic myeloid leukemia.** *Nature genetics* 2013, **45:**18-24.

223. Quesada V, Ramsay AJ, Lopez-Otin C: **Chronic lymphocytic leukemia with SF3B1 mutation.** *The New England journal of medicine* 2012, **366:**2530.

224. Wang L, Lawrence MS, Wan Y, Stojanov P, Sougnez C, Stevenson K, Werner L, Sivachenko A, DeLuca DS, Zhang L, et al: **SF3B1 and other novel cancer genes in chronic lymphocytic leukemia.** *The New England journal of medicine* 2011, **365:**2497-2506.

225. Wei X, Walia V, Lin JC, Teer JK, Prickett TD, Gartner J, Davis S, Stemke-Hale K, Davies MA, Gershenwald JE, et al: **Exome sequencing identifies GRIN2A as frequently mutated in melanoma.** *Nature genetics* 2011, **43:**442-446.

226. Davis RJ: **Signal transduction by the JNK group of MAP kinases.** *Cell* 2000, **103:**239-252.

227. Weston CR, Davis RJ: **The JNK signal transduction pathway.** *Current opinion in genetics & development* 2002, **12:**14-21.

228. Nikolaou K, Tsagaratou A, Eftychi C, Kollias G, Mosialos G, Talianidis I: **Inactivation of the deubiquitinase CYLD in hepatocytes causes apoptosis, inflammation, fibrosis, and cancer.** *Cancer Cell* 2012, **21:**738-750.

229. Das M, Garlick DS, Greiner DL, Davis RJ: **The role of JNK in the development of hepatocellular carcinoma.** *Genes & development* 2011, **25:**634-645.

230. Bennett BL, Sasaki DT, Murray BW, O'Leary EC, Sakata ST, Xu W, Leisten JC, Motiwala A, Pierce S, Satoh Y, et al: **SP600125, an anthrapyrazolone inhibitor of Jun N-terminal kinase.** *Proc Natl Acad Sci U S A* 2001, **98:**13681-13686.

231. Jiang DK, Sun J, Cao G, Liu Y, Lin D, Gao YZ, Ren WH, Long XD, Zhang H, Ma XP, et al: **Genetic variants in STAT4 and HLA-DQ genes confer risk of hepatitis B virus-related hepatocellular carcinoma.** *Nat Genet* 2012, **45:**72-75.

232. Jacquemin E: **Progressive familial intrahepatic cholestasis.** *Clinics and research in hepatology and gastroenterology* 2012, **36 Suppl 1:**S26-35.

233. Melino M, Gadd VL, Walker GV, Skoien R, Barrie HD, Jothimani D, Horsfall L, Jones A, Sweet MJ, Thomas GP, et al: **Macrophage secretory products induce an inflammatory phenotype in hepatocytes.** *World journal of gastroenterology : WJG* 2012, **18:**1732-1744.

234. Sun B, Karin M: **Inflammation and liver tumorigenesis.** *Frontiers of medicine* 2013, **7:**242-254.

235. Liu S, Gallo DJ, Green AM, Williams DL, Gong X, Shapiro RA, Gambotto AA, Humphris EL, Vodovotz Y, Billiar TR: **Role of toll-like receptors in changes in gene expression and NF-kappa B activation in mouse hepatocytes stimulated with lipopolysaccharide.** *Infection and immunity* 2002, **70:**3433-3442.

236. Machida K, Tsukamoto H, Mkrtchyan H, Duan L, Dynnyk A, Liu HM, Asahina K, Govindarajan S, Ray R, Ou JH, et al: **Toll-like receptor 4 mediates synergism between alcohol and HCV in hepatic oncogenesis involving stem cell marker Nanog.** *Proceedings of the National Academy of Sciences of the United States of America* 2009, **106:**1548-1553.

237. Schwabe RF, Seki E, Brenner DA: **Toll-like receptor signaling in the liver.** *Gastroenterology* 2006, **130:**1886-1900.

238. Matsumura T, Degawa T, Takii T, Hayashi H, Okamoto T, Inoue J, Onozaki K: **TRAF6-NF-kappaB pathway is essential for interleukin-1-induced TLR2 expression and its functional response to TLR2 ligand in murine hepatocytes.** *Immunology* 2003, **109:**127-136.

239. Li K, Chen Z, Kato N, Gale M, Jr., Lemon SM: **Distinct poly(I-C) and virus-activated signaling pathways leading to interferon-beta production in hepatocytes.** *The Journal of biological chemistry* 2005, **280:**16739-16747.

240. Yuhas Y, Berent E, Ashkenazi S: **Effect of rifampin on production of inflammatory mediators in HepG2 liver epithelial cells.** *Antimicrobial agents and chemotherapy* 2011, **55:**5541-5546.

241. Hogaboam CM, Bone-Larson CL, Steinhauser ML, Matsukawa A, Gosling J, Boring L, Charo IF, Simpson KJ, Lukacs NW, Kunkel SL: **Exaggerated hepatic injury due to acetaminophen challenge in mice lacking C-C chemokine receptor 2.** *The American journal of pathology* 2000, **156:**1245-1252.

242. Dong W, Simeonova PP, Gallucci R, Matheson J, Fannin R, Montuschi P, Flood L, Luster MI: **Cytokine expression in hepatocytes: role of oxidant stress.** *Journal of interferon & cytokine research : the official journal of the International Society for Interferon and Cytokine Research* 1998, **18:**629-638.

243. Sharrocks AD: **The ETS-domain transcription factor family.** *Nature reviews Molecular cell biology* 2001, **2:**827-837.

244. Hirschfield GM, Xie G, Lu E, Sun Y, Juran BD, Chellappa V, Coltescu C, Mason AL, Milkiewicz P, Myers RP, et al: **Association of primary biliary cirrhosis with variants in the CLEC16A, SOCS1, SPIB and SIAE immunomodulatory genes.** *Genes and immunity* 2012, **13:**328-335.

245. Mells GF, Floyd JA, Morley KI, Cordell HJ, Franklin CS, Shin SY, Heneghan MA, Neuberger JM, Donaldson PT, Day DB, et al: **Genome-wide association study identifies 12 new susceptibility loci for primary biliary cirrhosis.** *Nature genetics* 2011, **43:**329-332.

246. Liu X, Invernizzi P, Lu Y, Kosoy R, Bianchi I, Podda M, Xu C, Xie G, Macciardi F, Selmi C, et al: **Genome-wide meta-analyses identify three loci associated with primary biliary cirrhosis.** *Nature genetics* 2010, **42:**658-660.

247. Wada T, Joza N, Cheng HY, Sasaki T, Kozieradzki I, Bachmaier K, Katada T, Schreiber M, Wagner EF, Nishina H, Penninger JM: **MKK7 couples stress signalling to G2/M cell-cycle progression and cellular senescence.** *Nature cell biology* 2004, **6:**215-226.

248. Hui L, Bakiri L, Mairhorfer A, Schweifer N, Haslinger C, Kenner L, Komnenovic V, Scheuch H, Beug H, Wagner EF: **p38alpha suppresses normal and cancer cell proliferation by antagonizing the JNK-c-Jun pathway.** *Nat Genet* 2007, **39:**741-749.

249. Schramek D, Kotsinas A, Meixner A, Wada T, Elling U, Pospisilik JA, Neely GG, Zwick RH, Sigl V, Forni G, et al: **The stress kinase MKK7 couples oncogenic stress to p53 stability and tumor suppression.** *Nat Genet* 2011, **43:**212-219.

250. Eferl R, Wagner EF: **AP-1: a double-edged sword in tumorigenesis.** *Nature reviews Cancer* 2003, **3:**859-868.

251. Sakurai T, Maeda S, Chang L, Karin M: **Loss of hepatic NF-kappa B activity enhances chemical hepatocarcinogenesis through sustained c-Jun N-terminal kinase 1 activation.** *Proceedings of the National Academy of Sciences of the United States of America* 2006, **103:**10544-10551.

252. Hui L, Zatloukal K, Scheuch H, Stepniak E, Wagner EF: **Proliferation of human HCC cells and chemically induced mouse liver cancers requires JNK1-dependent p21 downregulation.** *J Clin Invest* 2008, **118:**3943-3953.

253. Chang Q, Zhang Y, Beezhold KJ, Bhatia D, Zhao H, Chen J, Castranova V, Shi X, Chen F: **Sustained JNK1 activation is associated with altered histone H3 methylations in human liver cancer.** *Journal of hepatology* 2009, **50:**323-333.

254. Min L, Ji Y, Bakiri L, Qiu Z, Cen J, Chen X, Chen L, Scheuch H, Zheng H, Qin L, et al: **Liver cancer initiation is controlled by AP-1 through SIRT6-dependent inhibition of survivin.** *Nat Cell Biol* 2012, **14:**1203-1211.

255. Ahn M, Witting SR, Ruiz R, Saxena R, Morral N: **Constitutive expression of short hairpin RNA in vivo triggers buildup of mature hairpin molecules.** *Hum Gene Ther* 2011, **22:**1483-1497.

256. McCaffrey AP, Meuse L, Pham TT, Conklin DS, Hannon GJ, Kay MA: **RNA interference in adult mice.** *Nature* 2002, **418:**38-39.

ACKNOWLEDGEMENTS

This work has been made possible thanks to many people.

My foremost gratitude goes to my supervisor, Gioacchino Natoli, who trusted in me and assigned me this exciting and ambitious project. His scientific expertise, dedication and fairness have been a great source of motivation during my graduate experience, and constitute a reference for the qualities I would like to embody in my future career (whatever path will it follow).
I will forever be thankful to Serena Ghisletti, who designed the project and led me throughout all the daily technical and conceptual challenges it presented. Her continuous support helped me becoming confident in my work. Most importantly, she taught me organization, patience and a pinch of diplomacy.
I would like to thank all the members of the GN group for helpful discussions, and in particular Paola Nicoli, for precious help with mouse colony maintenance and sample preparation and for being a great company in our long days at the animal facility, and Alberto Termanini, for carrying out the bioinformatic analyses included in the epigenetic part of the project and for his patience.
I am grateful to Francesca Ciccarelli (King's College, London, UK) and her group, in particular Fabio Iannelli (former member) and Shruti Sinha, for the design and the analysis of the experiments regarding the genetic alterations, and for scientific discussion of the work. It has been a great opportunity for me to work with and learn from them.
A special thank you to my co-supervisors, Bruno Amati at IEO and Piero Carninci at RIKEN (Yokohama, Japan), for critical discussion and support.
Furthermore, I thank Enrico Radaelli (KU Leuven Center for Human Genetics, Leuven, Belgium), for invaluable help with the histopathologic characterization of our specimens. I also thank Lorenzo D'Antiga and Aurelio Sonzogni (Hospital Papa Giovanni XXIII, Bergamo), Jamila Faivre and Marie Annick Buendia (INSERM, Villejuif, France), Ekkehard Sturm (University of Tuebingen, Germany) and Richard J. Thompson and Alexander S. Knisely (King's College Hospital, London, UK) for kindly donating the PFIC samples.
Many thanks to the NGS Unit at IFOM-IEO Campus, in particular to Luca Rotta, Salvatore Bianchi, for patient help with sample sequencing, and to Laura Tizzoni and Valentina Dall'Olio from the qPCR facility, who performed the CNV Taqman assay.

Last but not least, I would like to thank the people that supported this work by supporting me (which, in Italian, could sound both like "supportare" and "sopportare"…): Federica, Tiziano, Kiki, Giulia, Piwi, Elisa, Tommaso. They saw all the ups and downs of this PhD, and always stood at my side. A big thank you goes to all the Pro-Test crew, who shared my ideas and made me feel home.
Of course, all my love and gratitude to my family, who will not understand this because they can't speak English, but really made everything possible, and Stefano, the next Nobel Prize for Peace.

Grazie!